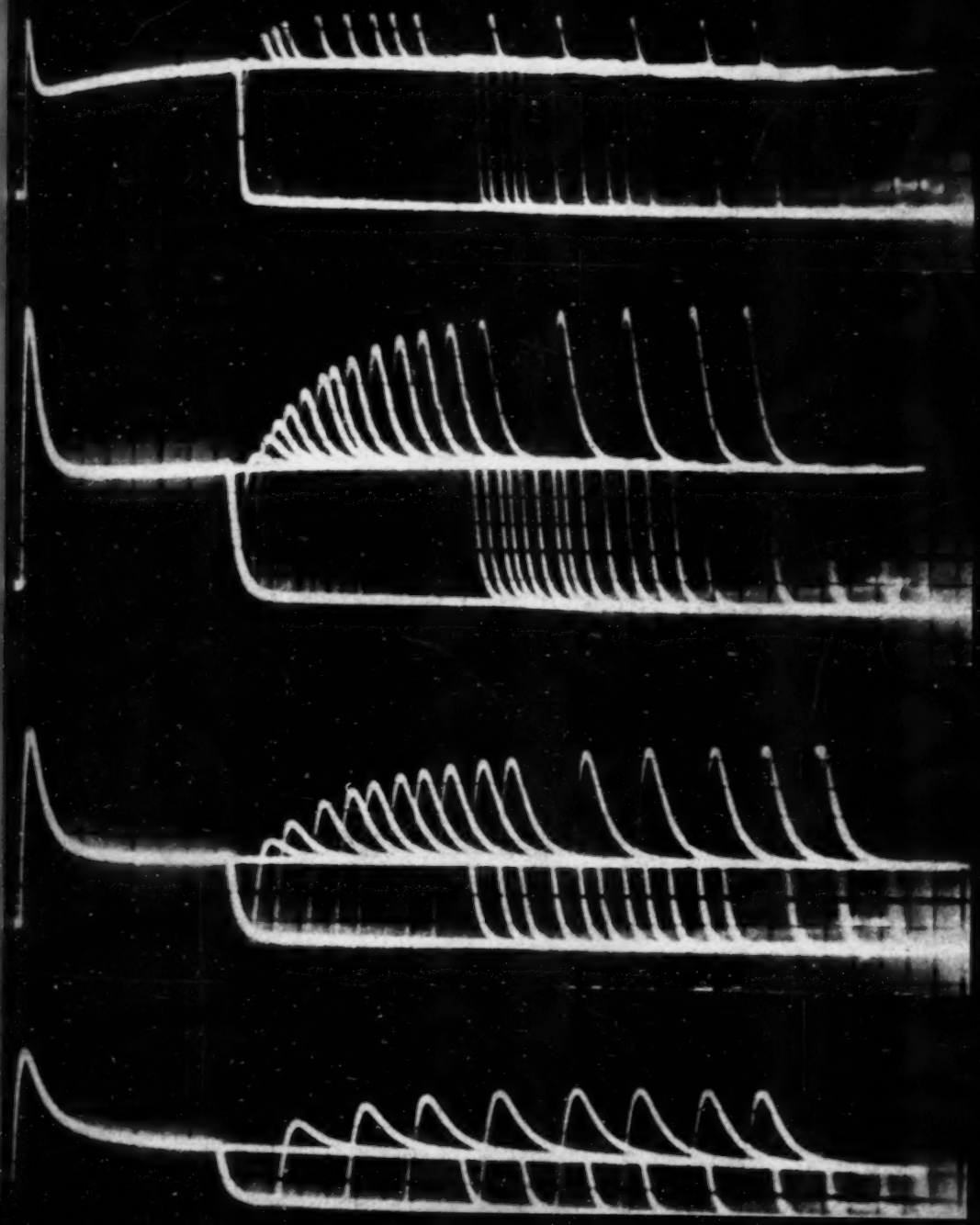
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 22, No. 2 FEBRUARY, 1951

In This Issue

	Page
Propagation Characteristics in a Coaxial Structure with Two Dielectrics — Alfredo Baños, Jr., David S. Saxon, and H. Gruen	117
Compressible Supersonic Flow in Jets under the Kármán-Tsien Pressure-Volume Relation — N. Coburn	124
Mechanical Determination of Biaxial Residual Stress in Sheet Materials — R. G. Treuting and W. T. Read, Jr.	130
Objective for Use in the Electron Microscopy of Ultra-Thin Sections — James Hillier	135
Leaky-Condenser Oxide Cathode Interface — A. Eisenstein	138
Some Measurements of Thermal Transpiration — S. Chu Liang	148
On the Directional Patterns of Polystyrene Rod Antennas — R. B. Watson	154
Asymptotic Approximation for the Normal Modes in Sound Channel Wave Propagation — N. A. Haskell	157
On the X-Ray Absorption Correction for Encased Diffractors in the Debye-Scherrer Technique — H. L. Ritter, R. L. Harris, and R. E. Wood	169
Dry Metallic Friction as a Function of Temperature Between 4.2°K and 600°K — I. Šimon, H. O. McMahon, and R. J. Bowen	177
Optimum Source Size for the Mach-Zehnder Interferometer — F. D. Bennett	184
Periodic Statistical Distortion of Unidirectionally Ordered Diffractors, with Application to Collagen — Richard S. Bear and Orvil E. A. Bolduan	191
Crystal Systems with Low Loss — Ruth Baumann Fry and William J. Fry	198
Density Measurements in Supersonic Flow by Means of the X-Ray Absorption Method — Eva M. Winkler	201
Deceleration and Ionizing Efficiency of Radar Meteors — D. W. R. McKinley	202
Latex Particle Size from X-Ray Diffraction Peaks — K. L. Yudowitch	214
Design of Molecular Pumps — Robert B. Jacobs	217
Hypersonic Flow at a Mach Number of 10 — I. D. V. Faro, T. R. Small, and F. K. Hill	220
Calculation of the Electrical Capacitance of a Cube — Daniel Kinseth Reitan and Thomas James Higgins	223
Properties of Evaporated Gold and Tungsten Oxides — P. G. Wilkinson	225
Letters to the Editor:	
Condensation of Water Vapor in the Shock Tube Below 150°K — P. Wegener and G. Lundquist	233
Observation of Initial Flow in a Fluid Obeying Darcy's Law by Radioactive-Tracer Techniques — F. T. Rogers, Jr., and L. E. Schilberg	233
Dynamic Method for Measuring the Specific Heat of Metals — Alan M. Nathan	234
Random Noise in Dielectric Materials — J. H. Mason	235
Cylindrical Cavity Filled with a dc Discharge — George W. Stuart, Jr., and Philip Rosen	236
Erratum: Original Authorship of Formulas Given in Clark and Deutsch's Paper — M. Reiner and G. W. Scott Blair	236
Index to Advertisers	xxiv

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Earle C. Gregg, *Assistant Editor*

Associate Editors

Howard H. Aiken
Lyle D. Borst
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the

American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
J. B. Fisk
G. P. Harnwell
George R. Harrison
W. V. Houston
Rudolph Kingslake
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
Francis G. Slack
John C. Steinberg

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Advertising Manager and Assistant Treasurer*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*
Kathryn Setze *Senior Accountant*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 22, Number 2 February, 1951

Propagation Characteristics in a Coaxial Structure with Two Dielectrics

ALFREDO BAÑOS, JR., DAVID S. SAXON, AND H. GRUEN*
University of California, Los Angeles, California
(Received March 21, 1950)

The propagation characteristics for circularly symmetric *TM* modes in a coaxial structure with two dielectrics offer many points of interest. For a given mode, the pertinent parameters are the ratio of the dielectric constants of the two media involved, the ratio of the two radii, and the operating frequency, expressed, say, in terms of the cut-off frequency for the *empty* guide. Among the results of special interest are the facts, previously pointed out by Frankel and by Bruck and Wicher, that the phase velocity for a given propagating mode may lie between the phase velocities corresponding to each of the two unbounded dielectrics and that, by proper choice of parameters, it is always possible to obtain a phase velocity of a preassigned value higher, of course, than the lower of the two velocities for the unbounded media. This suggests the possibility of using such structures in linear accelerators. The results of an extensive computational program are presented in the form of families of curves showing the dependence of the propagation constant and of the phase velocity on the parameters of interest. Some discussion of the power flow and field distribution is also given.

I. INTRODUCTION

IN the following there is presented the results of some extensive computations of the propagation constants for the two lowest circularly symmetric transverse magnetic modes in a circular wave guide filled with two concentric dielectrics (Fig. 1). The present authors first became interested in such a configuration in connection with the theoretical possibility of determining the complex dielectric constant of a gaseous discharge at microwave frequencies from a measurement of the propagation constant of waves in a circular guide containing a discharge along its axis. Thus, in Fig. 1, the outer region might be glass or some other suitable dielectric, the inner region, then, containing the discharge to be studied.

A second possible application of such wave guides, previously pointed out by Frankel[1] and by Bruck and Wicher,[2] lies in the fact that the phase velocity in the guide can be made less than its free space value, c, by an appropriate choice of dielectric constants and relative dimensions. Hence, such a system, with the central section evacuated, might be used as an accelerator tube in an electron linear accelerator, although, presumably, the higher power level required would place stringent requirements on the dielectric with respect to its breakdown properties.[3] In any case, the problem has an

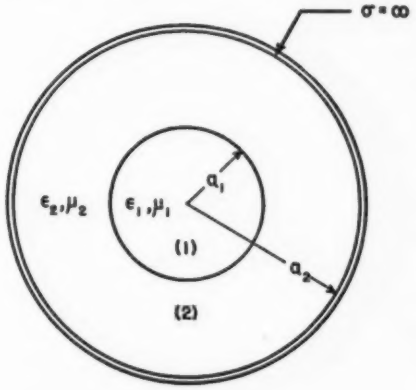


FIG. 1.

* Presently also with the National Bureau of Standards, Institute for Numerical Analysis, Los Angeles, California.
[1] Sidney Frankel, J. Appl. Phys. 18, 650–55 (1947).
[2] G. G. Bruck and E. R. Wicher, J. Appl. Phys. 18, 766–9 (1947).
[3] More recently, R. B. R. Shersbie-Harvie, Nature 162, 890 (1948), has proposed an interesting new form of dielectric loaded wave guide for linear accelerators in which the outer dielectric is made to appear anisotropic, with a dielectric constant much smaller in the longitudinal direction than in the transverse or radial direction, the required anisotropy being obtained by means of a stack of thin isotropic laminas of suitable thickness and properly spaced. It is claimed that this arrangement materially reduces the metallic losses of the outer envelope. In this investigation we have considered only isotropic dielectrics, but the extension to anisotropic outer dielectrics is straightforward.

intrinsic interest as a rather different kind of wave guide problem.

Because of the evinced interest in the linear accelerator applications of a coaxial structure with two dielectrics, we are presenting the results in the form of curves which give essentially, for a wide choice of parameters, the frequency dependence of the guide phase velocity for the lowest modes. We have also discussed the power flow and the field distribution. Although these latter results were contained essentially in Frankel's paper, we feel that the presentation of the curves, which greatly facilitates the solution of numerous problems, well warranted carrying this investigation a step further.

II. SOLUTION OF THE BOUNDARY VALUE PROBLEM

As is well known,[4] the entire electromagnetic field in cylindrical coordinates (r, θ, z), can be expressed in terms of two scalar functions, one of which generates a field whose magnetic intensity vector is purely transverse ($H_z=0$), while the other generates a field whose electric intensity vector is purely transverse ($E_z=0$). The former is called an electric type or transverse magnetic (TM) field, whereas the latter is designated as a magnetic type or transverse electric (TE) field.

In particular, the circularly symmetric TM modes with which we shall be concerned are generated from a scalar function of the type

$$\psi(r, z, t)=Z_0(\gamma r)e^{\pm ihz-i\omega t}, \tag{2.01}$$

through the expressions

$$\mathbf{E}=\nabla\times\nabla\times(e_z\psi); \quad \mathbf{H}=-i\omega\epsilon\nabla\times(e_z\psi), \tag{2.02}$$

where e_z denotes a unit vector in the direction of the positive z axis. In the above, $Z_0(\gamma r)$ is an arbitrary cylinder function of order zero, while the transverse and longitudinal wave numbers, γ and h respectively, are related to the assumed angular frequency ω by the condition

$$k^2=\left(\frac{2\pi}{\lambda}\right)^2=\omega^2\mu\epsilon=h^2+\gamma^2, \tag{2.03}$$

where ϵ and μ are the electric and magnetic inductive capacities of the unbounded medium. It should be mentioned that rationalized mks units will be used throughout, so that in vacuum, for example, we have

$$1/(\mu_0\epsilon_0)^{\frac{1}{2}}=c\simeq 3\times 10^8 \text{ meters/sec},$$
$$(\mu_0/\epsilon_0)^{\frac{1}{2}}\simeq 120\pi=377 \text{ ohms}.$$

In general, if the fields are confined within some cylindrical structure, the boundary conditions at its surface determine a set of transverse wave numbers γ, whence from Eq. (2.03) h is determined, for each γ, in terms of the wave number k or, equally well, in terms of the angular frequency ω. If k is greater than a particular γ, then h is real and the corresponding field can propagate along the axis of the structure. However, if k is less than γ, then h is imaginary and the field is not propagated. Thus, γ is often called the cut-off wave number for the mode in question. Since γ is real for all simple wave guide structures, it is immediately apparent from Eq. (2.03) that, for propagation, $h<k$ and, hence, the phase velocity of waves guided by such configurations is always larger than the phase velocity for plane homogeneous waves in the unbounded medium.

To proceed with the problem of the coaxial structure of Fig. 1, consider first the inner dielectric of radius a_1 and inductive capacity ϵ_1. The particular scalar function $\psi_1(r, z, t)$ from which the fields in this region are to be derived is, evidently, of the form

$$\psi_1(r, z, t)=A_1J_0(\gamma_1 r)e^{\pm ihz-i\omega t}, \tag{2.04}$$

where A_1 is an arbitrary constant, to be determined later from the boundary conditions and the average power flow, and γ_1 is the corresponding transverse wave number. Here, we have chosen, for the cylinder function appearing in Eq. (2.01), the Bessel function $J_0(\gamma_1 r)$, since the fields must remain finite on the axis of the structure.† The fields in the inner region are, then, given in terms of ψ_1 by Eq. (2.02).

In the outer dielectric, of external radius a_2 and inductive capacity ϵ_2, the field is somewhat more complicated. However, assuming that the conductivity of the metallic boundary of the structure is infinite, it is seen from Eq. (2.02) that the function ψ_2, which generates the field in this region, must vanish on the boundary $r=a_2$, whence, we can write at once

$$\psi_2(r, z, t)=A_2[J_0(\gamma_2 r)N_0(\gamma_2 a_2) - N_0(\gamma_2 r)J_0(\gamma_2 a_2)]e^{\pm ihz-i\omega t}, \tag{2.05}$$

where γ_2 is the transverse wave number corresponding to this region and A_2 is another constant which, like A_1 is to be determined from the boundary conditions at $r=a_1$ and from a knowledge of the power level in the structure. As before, the fields in this region are determined by using Eq. (2.02).

It is observed that different transverse wave numbers γ_1 and γ_2 appear in Eqs. (2.04) and (2.05) although the propagation constant h is the same in both cases. The reason for this is, of course, that h must be the same if the boundary conditions at the interface between the two dielectrics are to be satisfied for all values of z and t, whence from Eq. (2.03), γ must be different in the two media and, indeed, we must have

$$\gamma_1^2=k_1^2-h^2=\omega^2\mu_0\epsilon_1-h^2;$$
$$\gamma_2^2=k_2^2-h^2=\omega^2\mu_0\epsilon_2-h^2, \tag{2.06}$$

by means of which γ_1 can be expressed in terms of γ_2 or vice versa. Here, we have written μ_0 for the magnetic

[4] See, for example, J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Chapter VI.

† The Neumann function $N_0(\gamma_1 r)$ becomes logarithmically infinite on the axis, corresponding to the existence of a line source, and, hence, it must be excluded in this region.

inductive capacity of both media assuming that both dielectrics are non-magnetic.

The value of the propagation constant h is now determined by demanding that the tangential field components E_z, H_θ be continuous upon crossing the interface at $r=a_1$. Thus, we find for E_z

$$\gamma_1^2 A_1 J_0(\gamma_1 a_1) = \gamma_2^2 A_2 [J_0(\gamma_2 a_1) N_0(\gamma_2 a_2) - N_0(\gamma_2 a_1) J_0(\gamma_2 a_2)], \quad (2.07)$$

and for H_θ

$$k_1^2 \gamma_1 A_1 J_1(\gamma_1 a_1) = k_2^2 \gamma_2 A_2 [J_1(\gamma_2 a_1) N_0(\gamma_2 a_2) - N_1(\gamma_2 a_1) J_0(\gamma_2 a_2)], \quad (2.08)$$

where we have made use of the fact that $\omega\epsilon = k^2/\omega\mu$ and that $\mu=\mu_0$ for both dielectrics. Eliminating A_1 and A_2 from these equations yields the highly transcendental equation for h

$$\frac{\gamma_1 J_0(\gamma_1 a_1)}{k_1^2 J_1(\gamma_1 a_1)} = \frac{\gamma_2 [J_0(\gamma_2 a_1) N_0(\gamma_2 a_2) - N_0(\gamma_2 a_1) J_0(\gamma_2 a_2)]}{k_2^2 [J_1(\gamma_2 a_1) N_0(\gamma_2 a_2) - N_1(\gamma_2 a_1) J_0(\gamma_2 a_2)]}, \quad (2.09)$$

while the ratio A_1/A_2 is obtained from Eqs. (2.07) or (2.08). Equation (2.09), regarded as an equation in h through Eq. (2.06), can only be solved numerically or graphically. The method actually employed for the determination of the roots of Eq. (2.09) is described in Sec. IV.

III. POWER FLOW

Let us now examine the average power flow in such a coaxial structure with particular emphasis on the relative amounts flowing in each section. Thus, making use of Poynting's vector, it is easily seen that the average power flow down the guide is given by

$$P = \frac{1}{2} \int Re[E_r H_\theta^*] r d\theta dr, \quad (3.01)$$

where the integral extends over the cross section of the guide.

With the aid of the indefinite integral[5]

$$\int^r Z_1(\gamma r) Z_1^*(\gamma r) r dr = \tfrac{1}{2} r^2 \left[|Z_0(\gamma r)|^2 + |Z_1(\gamma r)|^2 - 2Re\left(\frac{Z_0^*(\gamma r) Z_1(\gamma r)}{\gamma r} \right) \right], \quad (3.02)$$

and of Eqs. (2.02), (2.04), (2.05), (2.07), (2.08), and (2.09), it is established after considerable manipulation that the total average power P flowing along the coaxial structure can be expressed as‡

$$P = \tfrac{1}{2} |E_z^0|^2 \pi a_1^2 \frac{h k_1}{|\gamma_1|^2} \left(\frac{\epsilon_1}{\mu_0} \right)^{\frac{1}{2}} J_0^2(\gamma_1 a_1) [Q_1 + Q_2]. \quad (3.03)$$

In the above, E_z^0 is the peak value of the z component of the electric field at the center of the guide and is expressible in terms of A_1 by

$$E_z^0 = \gamma_1^2 A_1, \quad (3.04)$$

and the dimensionless quantities Q_1 and Q_2, representing essentially the normalized power flow in the inner and outer regions, respectively, are given by

$$Q_1 = 1 + \left| \frac{J_1(\gamma_1 a_1)}{J_0(\gamma_1 a_1)} \right|^2 - \frac{2 J_1(\gamma_1 a_1)}{\gamma_1 a_1 J_0(\gamma_1 a_1)}, \quad (3.05)$$

and

$$Q_2 = \frac{k_2^2}{k_1^2} \left| \frac{2\gamma_1}{\pi \gamma_2^2 a_1 C_0(\gamma_2 a_1)} \right|^2 - \frac{k_2^2}{k_1^2} \left| \frac{\gamma_1}{\gamma_2} \right|^2 - \frac{k_1^2}{k_2^2} \left| \frac{J_1(\gamma_1 a_1)}{J_0(\gamma_1 a_1)} \right|^2 + \left| \frac{\gamma_1}{\gamma_2} \right|^2 \frac{2 J_1(\gamma_1 a_1)}{\gamma_1 a_1 J_0(\gamma_1 a_1)}, \quad (3.06)$$

where

$$C_0(\gamma_2 a_1) = J_0(\gamma_2 a_1) N_0(\gamma_2 a_2) - N_0(\gamma_2 a_1) J_0(\gamma_2 a_2). \quad (3.07)$$

It should be mentioned, perhaps, that some care must be taken in the deduction of Eq. (3.03) since either γ_1 or γ_2 may be purely imaginary, depending upon the choice of parameters, as will be shown later.

IV. GRAPHICAL SOLUTION OF THE TRANSCENDENTAL EQUATION

Before describing the graphical method for the determination of a number of roots of the transcendental equation Eq. (2.09), it will be convenient to introduce the symbol η to denote the ratio of outer to inner radius and the symbol κ to denote the ratio of the inner to the outer dielectric constant; thus,

$$\eta = \frac{a_2}{a_1}; \quad \kappa = \frac{\kappa_1}{\kappa_2} = \frac{\epsilon_1}{\epsilon_2} = \frac{k_1^2}{k_2^2}, \quad (4.01)$$

where, it will be recalled, $k_1^2 = \mu_0 \epsilon_0 \kappa_1 \omega^2$ and $k_2^2 = \mu_0 \epsilon_0 \kappa_2 \omega^2$, since we have assumed the two dielectrics to be non-magnetic. Furthermore, to simplify the treatment that follows, it will be convenient to introduce the dimensionless variables

$$\left. \begin{aligned} u &= k_2 a_1 \\ v &= \gamma_2 a_1 \\ w &= h a_1 \\ x &= \gamma_1 a_1 \end{aligned} \right\} \quad (4.02)$$

among which there exist, by virtue of Eq. (2.06) which

[5] G. N. Watson, *Theory of Bessel Functions* (The Macmillan Company, New York, 1944), second edition, p. 134, making use of Eq. (10).

‡ This result was essentially established earlier by Frankel, (reference 1).

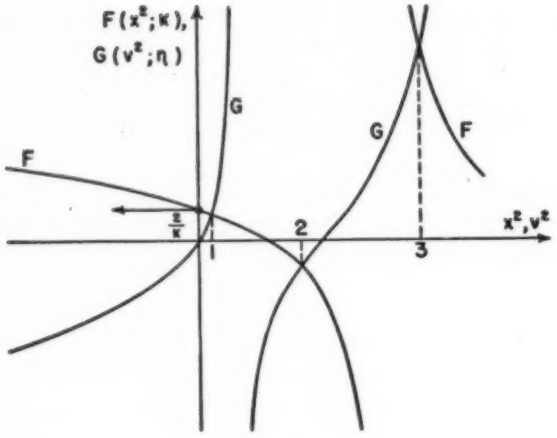


FIG. 2.

implies $h^2=k_1^2-\gamma_1^2=k_2^2-\gamma_2^2$, the relations

$$x^2=\kappa u^2-w^2; \quad v^2=u^2-w^2. \tag{4.03}$$

In terms of these new variables, our transcendental equation (2.09) may now be rewritten as

$$F(x^2;\kappa)=G(v^2;\eta), \tag{4.04}$$

where

$$F(x^2;\kappa)=\frac{xJ_0(x)}{\kappa J_1(x)} \quad \text{for} \quad x^2\gtrless 0,$$
$$=\frac{sI_0(s)}{\kappa I_1(s)} \quad \text{for} \quad x=is,\ x^2\lesseqgtr 0; \tag{4.05}$$

and§

$$G(v^2;\eta)=v\frac{J_0(v)N_0(\eta v)-N_0(v)J_0(\eta v)}{J_1(v)N_0(\eta v)-N_1(v)J_0(\eta v)} \quad \text{for} \quad v^2\gtrless 0,$$
$$=t\frac{I_0(t)K_0(\eta t)-K_0(t)I_0(\eta t)}{I_1(t)K_0(\eta t)+K_1(t)I_0(\eta t)} \tag{4.06}$$
$$\text{for} \quad v=it,\ v^2\lesseqgtr 0.$$

The functions F and G, which are real for the specified range of their respective arguments, are depicted graphically in Fig. 2 where the abscissas, labelled 1, 2, 3, $\cdots$, of the intersections between these two curves yield, through Eq. (4.03), the roots of the transcendental equation (4.04). In fact, if we regard the applied frequency as an independent parameter, we see that $u=k_2a_1=\omega a_1(\mu_0\epsilon_0\kappa_2)^{\frac{1}{2}}$ is proportional to ω and may, thus, be regarded as a (normalized) measure of the frequency. Furthermore, eliminating $w=ha_1$ from Eq. (4.03), we may write

$$x^2=v^2+(\kappa-1)u^2, \tag{4.07}$$

which exhibits x^2 in terms of v^2 for prescribed values of u (or of the frequency), and of the ratio κ. When $u=0$ (or when $\kappa=1$, whatever the value of u), we have from Eq. (4.07) $x^2=v^2$, which is the situation depicted in Fig. 2, as far as the function $F(x^2;\kappa)$ is concerned. For fixed κ and any other value of u, the relation Eq. (4.07) states that the curve $F(v^2;\kappa)$ is to be obtained from the curve $F(x^2;\kappa)$, shown in Fig. 2, by translating the latter to the left ($\kappa>1$) as indicated by the arrow in the figure, or to the right ($\kappa<1$), by an amount $|\kappa-1|u^2$ as measured on the v^2 axis. For given η we have a definite curve $G(v^2;\eta)$ which stays fixed and, as we vary the frequency for given κ, we are to translate laterally the curve $F(x^2;\kappa)$ in accordance with Eq. (4.07). For every given situation, we are to read off the intersection abscissas v_1^2, v_2^2, etc., and inserting these values in Eq. (4.03), we obtain, finally,

$$w^2=u^2-v_n^2, \quad n=1, 2, \cdots. \tag{4.08}$$

Since $w=ha_1$, this procedure yields the values of the propagation constant h for given κ, η, and ω.

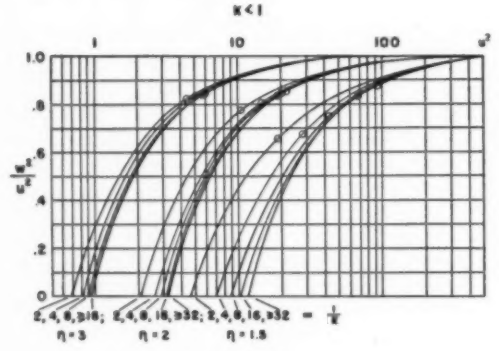


FIG. 3.

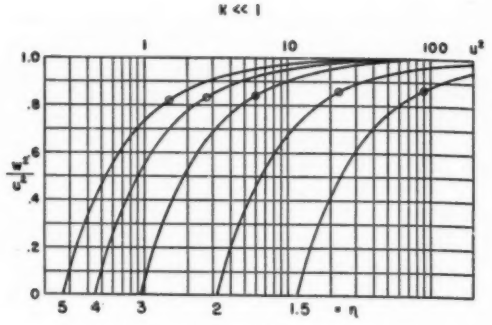


FIG. 4.

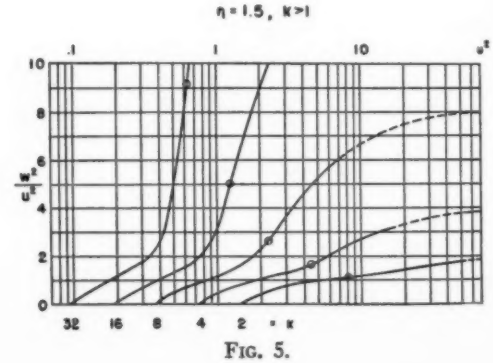


FIG. 5.

§ Here $K_n(t)$ denotes Macdonald's Bessel function of the second kind for purely imaginary argument. See G. N. Watson, reference 5, p. 78.

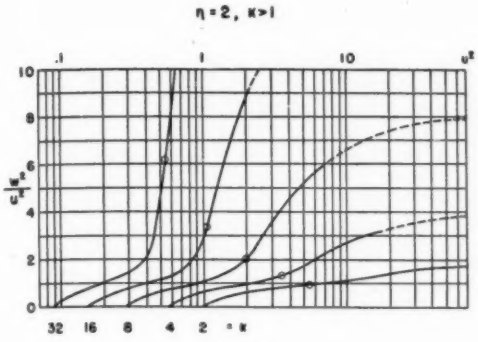


FIG. 6.

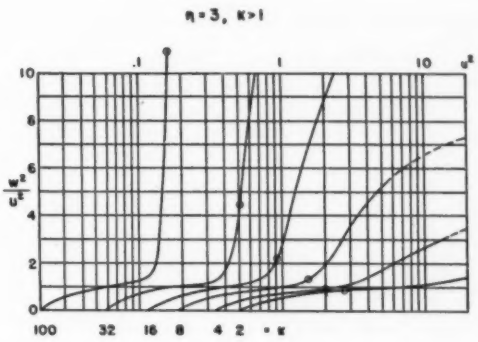


FIG. 7.

The adopted graphical procedure consisted, then, of preparing three large scale graphs of the function $G(v^2;\eta)$, corresponding to $\eta=1.5$, 2, 3, and, drawn to the same scale, two families of curves $F(x^2;\kappa)$ for $\kappa<1$ and $\kappa>1$, respectively. It proved convenient to choose $\kappa=2^n$, $n=1,\cdots,5$, for $\kappa>1$, and the reciprocal values $\kappa=1/2^n$ for $\kappa<1$. Once in possession of these plots, the determination of the intersection abscissas proceeded as outlined above. The investigation was limited, in each case, to the first two modes of operation, since practical interest centers around the dominant mode and its frequency range of operation.

Further, as stated in the introduction, one special feature of this coaxial structure is that any pre-assigned value of the phase velocity $V_p=\omega/h$ can be obtained, within the limits determined below, by a suitable choice of the parameters κ and η. Hence, the ratio

$$w^2/u^2=h^2/k_2^2=c^2/\kappa_2 V_p^2, \tag{4.09}$$

where c is the velocity of light in free space, may be of some interest. In consequence, we have prepared the results of these studies in the form of graphs exhibiting the dimensionless ratio Eq. (4.09), for each particular choice of parameters, as a function of the independent variable

$$u^2=\mu_0\epsilon_0\kappa_2 a_1^2\omega^2, \tag{4.10}$$

which is, for a given structure, proportional to the square of the applied angular frequency.

For $\kappa<1$, Fig. 3 gives, in semilogarithmic coordinates, three families of curves of the ratio $(w/u)^2$ *vs* u^2, corresponding to the lowest circularly symmetric *TM* mode for each of the chosen values of η and κ which are indicated on the graph. It is seen from Eq. (4.04) that, for the lowest mode,

$$w^2/u^2=1-(v_1^2/u^2), \tag{4.11}$$

where v_1^2 is the intersection abscissa labelled 1 in Fig. 2. Furthermore, it is clear from Eq. (4.07) and from Fig. 2 that, as $u^2\to\infty$ for $\kappa<1$, the intersection abscissa v_1^2 tends asymptotically to the first zero in the denominator of $G(v^2;\eta)$ as given in Eq. (4.06). Hence, as $u^2\to\infty$, the ratio $(w/u)^2\to 1$ for all values of $\kappa<1$, as indicated in Fig. 3. In this figure, as well as in Figs. 4 through 7, the abscissas corresponding to $w=0$ give, for each curve, a measure of the cut-off frequency of the corresponding lowest mode, whereas the abscissas of the small circles correspond similarly to the cut-off frequencies of the next higher mode. In this manner, it is possible to establish at a glance for each case the frequency range for lowest single mode operation. These values of u^2 are given in Table I, where u_1^2 corresponds to the lowest mode and u_2^2 to the next higher mode.

Figure 4 gives, for $\kappa\ll 1$, the asymptotic form of the curves in Fig. 3 for each value of η. That this asymptotic behavior exists in this case is readily seen from Fig. 2, where, as $\kappa\to 0$, the intercept of the curve $F(x^2;\kappa)$ on the axis of ordinates grows without limit implying that the intersection abscissas $v_n^2(n=1,2,\cdots)$, tend, for any value of u^2, to asymptotic values corresponding to the successive zeros in the denominator of $F(v^2;\eta)$; that is, as $\kappa\to 0$, the values of v_n tend asymptotically to the successive roots of

$$J_1(v)N_0(\eta v)-N_1(v)J_0(\eta v)=0. \tag{4.12}$$

Accordingly, for $\kappa\ll 1$, say $\kappa\lesssim 0.02$, the curves in Fig. 4 are given by the equation

$$w^2/u^2=1-[v_1^2(\eta)/u^2], \tag{4.13}$$

where $v_1(\eta)$ is the first root of Eq. (4.12), and similarly for the higher modes. The first two roots of Eq. (4.12) were computed by numerical methods for several values of η and are given in Table II.

TABLE I. Cut-off values for first two modes.

κ	$\eta=1.5$		$\eta=2.0$		$\eta=3.0$	
	u_1^2	u_2^2	u_1^2	u_2^2	u_1^2	u_2^2
$\frac{1}{32}$	11.59	90.65	3.19	23.06	0.935	5.94
$\frac{1}{16}$	10.20	67.30	3.07	21.40	0.935	5.78
$\frac{1}{8}$	8.90	44.00	2.86	19.35	0.900	5.50
$\frac{1}{4}$	7.00	28.60	2.57	15.20	0.810	5.09
$\frac{1}{2}$	4.63	19.60	2.11	10.60	0.690	4.48
1	2.57	13.52	1.44	7.62	0.643	3.38
2	1.400	8.17	0.960	5.52	0.500	2.63
4	0.720	4.47	0.550	3.54	0.323	1.98
8	0.365	2.33	0.283	1.99	0.183	1.50
16	0.196	1.23	0.149	1.06	0.110	0.91
32	0.096	0.63	0.084	0.55	0.058	0.51
100	—	—	—	—	0.020	0.16

TABLE II. First two roots of $J_1(v)N_0(\eta v)-N_1(v)J_0(\eta v)=0$.

η	1	1.5	2	3	4	5
v_1	∞	3.404	1.787	0.967	0.668	0.516
v_2	∞	9.521	4.802	2.437	1.656	1.248

For $\kappa>1$, Figs. 5–7 give, for each of the indicated values of η, the curves of $(w/u)^2$ *vs* u^2 corresponding to the lowest circularly symmetric *TM* mode. Here, each curve is labelled according to the value of κ to which it corresponds, and, as seen from the extrapolated dotted portions shown in the graphs, each curve tends asymptotically as $u^2\to\infty$ to the corresponding value of $\kappa>1$. This asymptotic behavior is readily established from Fig. 2 by observing that, as $u^2\to\infty$, the intersection abscissas $v_n^2(n=1, 2, \cdots)$, become, according to Eq. (4.07), approximately

$$v_n^2\approx x_n^2-(\kappa-1)u^2, \tag{4.14}$$

where $x_n(n=1, 2, \cdots)$, are the roots of $J_1(x)=0$, which appears in the denominator of the function $F(x^2;\kappa)$ as given in Eq. (4.05). Substituting Eq. (4.14) into Eq. (4.08), we readily obtain, as $u^2\to\infty$,

$$w^2/u^2\approx\kappa-(x_n^2/u^2), \tag{4.15}$$

from which ensues the asymptotic behavior claimed above. In these figures, as before, the intercepts on the axis of abscissas correspond to the lowest mode cut-off frequencies, and the abscissas of the small circles correspond similarly to the next higher mode. The corresponding numerical values are given in Table I.

V. NUMERICAL EXAMPLES

It is seen at once from the parameter relationships Eq. (4.03), that three distinct cases arise in practice under which the transcendental equation and the field components have different expressions:

(1). The parameters u, v, w, and x are all real. In this case, as seen from Eq. (4.09), $(V_p/c)^2\geqq 1/\kappa_2$ for $\kappa<1$, or $(V_p/c)^2\geqq 1/\kappa_1$ for $\kappa>1$. Thus, here, the ratio of the phase velocity to the velocity of light in free space is seen to have a lower limit (<1) determined by the larger of the two dielectric constants.

(2). The parameters u, v, and w are real, whereas $x=is$ is pure imaginary. In this case, $\kappa<1$, the phase velocity is bracketed between the limits $1/\kappa_2<(V_p/c)^2<1/\kappa_1$.

(3). The parameters u, v, and x are real, whereas $v=it$ is pure imaginary. Here $\kappa>1$ and the phase velocity lies within the limits $1/\kappa_1<(V_p/c)^2<1/\kappa_2$.

Two numerical examples will now be given to illustrate the use of the above results. In the first example, which may possibly apply to the linear accelerator, let the inner dielectric be free space and let us require that the operating phase velocity be $V_p=0.8c$, whence, from Eq. (4.09), $(u/w)^2=0.64\kappa_2$. Choose $\kappa_1=1$, $\kappa_2=100$, and $\eta=3$. Here, $\kappa=0.01\ll 1$, and it is immediately obvious that the asymptotic form Eq. (4.13) is applicable in this case; i.e.,

$$\frac{w^2}{u^2}=0.0156=1-\frac{(0.967)^2}{u^2},$$

from which, making use of Eq. (4.03), we finally obtain

$$u=0.975;\quad v=0.967;\quad w=0.124;\quad x=is,\quad s=0.0731,$$

and, from Eq. (4.10), we have $\omega^2a_1^2=8.55\times 10^{14}$, from which the inner radius (in meters) is determined for a given operating angular frequency.

This is an example of case (2) listed above and, making use of the numerical values given, it is now possible to compute, in terms of the parameters $|E_z^0|$ and a_1, the

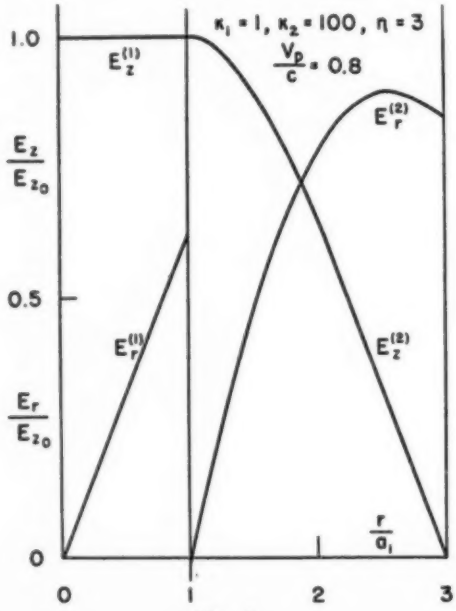


FIG. 8.

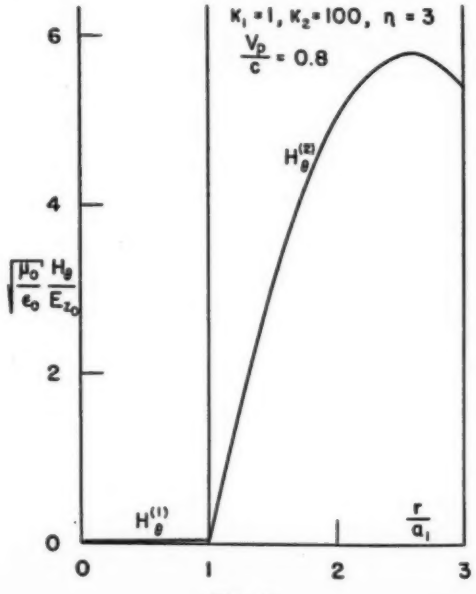


FIG. 9.

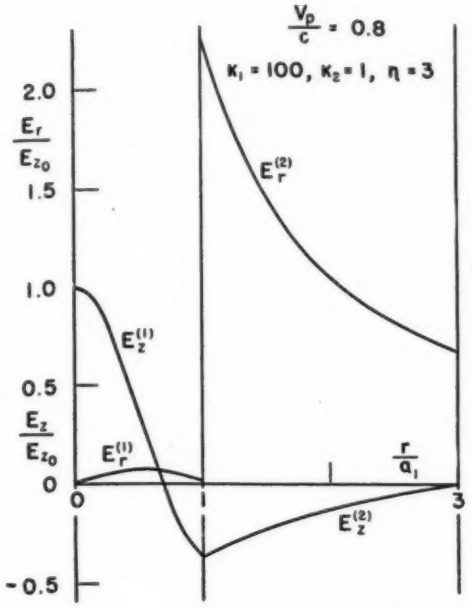


FIG. 10.

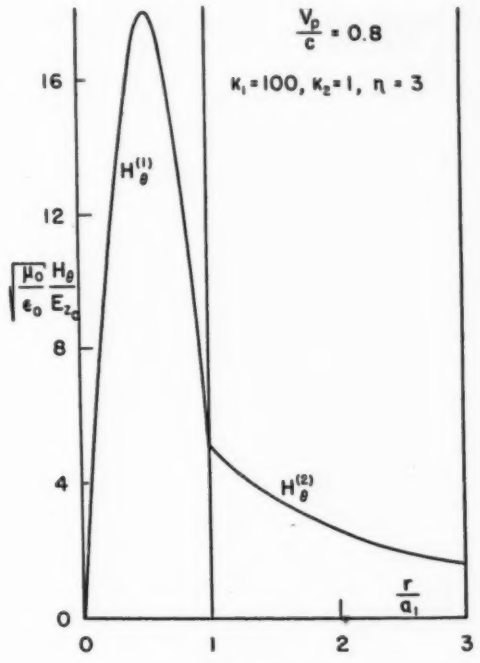


FIG. 11.

distribution of the field components and the relative power flow in this coaxial structure. Figures 8 and 9 give, for this case, the distribution of the electric and magnetic intensity components E_z, E_r, and H_θ, expressed as convenient dimensionless ratios *qua* function of r/a_1. These curves were computed through Eq. (2.02), after inserting the proper form of the Bessel and cylinder functions appropriate to this case. The total power flow, as computed from Eq. (3.03), is given by

$$P = 0.0156 |E_z^0|^2 a_1^2 \text{ watts},$$

whereas the power ratio becomes $Q_1/Q_2 = 0.0038$ or about 0.4 percent. Thus, it is seen that, in this case, most of the power flows through the outer dielectric.

As a second example, let us consider the converse proposition in which now $\kappa_1 = 100$, $\kappa_2 = 1$, and $\eta = 3$. Again, we shall demand that $V_p = 0.8c$, whence, here, $(u/w)^2 = 0.64$. This time $\kappa = 100$, and the asymptotic forms cannot be used; thus, we must read the value of u^2 off the corresponding curve ($\kappa = 100$) in Fig. 7. In this manner, we obtain, proceeding as before,

$$u = 0.349; \quad w = 0.437; \quad x = 3.46; \quad v = il, \quad t = 0.263,$$

and, again from Eq. (4.10), we have $\omega^2 a_1^2 = 1.10 \times 10^{16}$. The total power flow is now given by

$$P = 0.0344 |E_z^0|^2 a_1^2 \text{ watts}$$

with the power ratio $Q_1/Q_2 = 0.031$ or about 3 percent. Again, we see that most of the power flows in the outer space.

This is an example of case (3) listed above and, as before, we can now compute the space distribution of the field components. The results are given in Figs. 10 and 11. It is noteworthy that, in both of these examples, there exist two important relations; namely,

$$\omega^2 a_1^2 = C_1 \quad \text{and} \quad P/|E_z^0|^2 a_1^2 = C_2,$$

where C_1 and C_2 are constants fully determined, for a given coaxial structure, by κ_1, κ_2, η, and V_p/c. Obviously, many other relations can be established depending upon the type of operation required. Thus, the numerical examples treated have served merely to illustrate one class of problems that can now be solved with the aid of the curves and asymptotic relations established in this paper. It is hoped that the results presented here have made amenable to solution many other problems related to the propagation characteristics of a coaxial structure with two dielectrics.

Compressible Supersonic Flow in Jets under the Kármán-Tsien Pressure-Volume Relation

N. COBURN
University of Michigan, Ann Arbor, Michigan
(Received January 25, 1950)

The two-dimensional supersonic irrotational flow of a gas in a jet is studied by use of the Kármán-Tsien pressure-volume law. There are two limitations to such a study: (1) since the fluid flow is not continued from the subsonic range, arbitrary boundary conditions must be prescribed; (2) use of the Kármán-Tsien pressure-volume relation implies a restriction on the permissible range of pressure, density, and velocity. On the other hand, use of the Kármán-Tsien law furnishes several advantages: (1) the velocity potential and stream function satisfy the wave equation in the hodograph plane and hence these functions can be easily determined; (2) the mappings between the physical and hodograph planes may be completely characterized and studied in detail. This gain in information should be valuable in the qualitative understanding of phenomena as well as in obtaining first approximations to quantitative solutions. In the case of jets, with free stream lines as boundaries, it is shown that two functions *possessing certain desired properties* completely determine the Kármán-Tsien flow. Further, the phenomenon of the periodic recurrence of the free stream jet boundary is explained *by a folding property of the map of the flow in the hodograph plane.*

1. INTRODUCTION

OUR problem will be to develop a simplified theory for the two-dimensional, irrotational, supersonic flow of a compressible gas in a jet. Owing to the assumption of the validity of the Kármán-Tsien pressure-volume relation, the relation between the maps of the flow in the physical, hodograph, and stream function-velocity potential planes can be explicitly exhibited.

In the physical (x, y) plane, the velocity potential, ϕ, and the stream function, ψ, satisfy a nonlinear second-order partial differential equation. However, in the hodograph (ω, θ) plane (where ω is the angle between the characteristics in the physical plane and θ is the angle between the velocity vector and the x-axis of the physical plane), these functions satisfy a linear second-order partial differential equation. For the adiabatic pressure-volume relation, this equation has variable coefficients. In spite of the fact that this equation can not be explicitly integrated, Frankl[1] has discussed various existence theorems. If the adiabatic pressure-volume curve is approximated by use of the tangent line to this curve at an appropriately chosen point (this is the Kármán-Tsien pressure-volume relation), the above mentioned linear differential equation becomes the wave equation and explicit solutions may be found. When boundary value data are given along the line $\omega=\omega_1$ (this line can be taken arbitrarily close to the sonic line $\omega=\omega_*$), the solution of the wave equation depends upon the region of integration in the (ω, θ) plane. By examining the maps of the flow in the (ω, θ), (ϕ, ψ), and (x, y) planes, the proper integration regions in the (ω, θ) plane are determined. Thus, the velocity potential and stream function can be determined in the (ω, θ) plane and finally these functions can be mapped into the physical (x, y) plane.

Our results are the following: (1) an analytic method, which depends upon a knowledge of two arbitrary functions, for treating fairly uniform supersonic flows in jets with free stream lines is developed; (2) the phenomenon of the periodic recurrence of the free stream jet boundary is explained by a folding property of the map of the flow in the hodograph plane.

Before concluding this introductory section, we shall list some of the formulas[2] used in our previous general study of the Kármán-Tsien pressure-volume relation. If we replace the usual adiabatic gas law by a tangent line drawn at the point $(1/\rho_1, p_1)$ in the pressure-density diagram, one finds that

$$p_1-p=a_1^2\rho_1^2[(1/\rho)-(1/\rho_1)], \tag{1.1}$$

where ρ is density, p is the pressure, a is the local sound speed. From (1.1) and the Bernoulli relation, it follows that

$$a^2\rho^2=k^2, \quad w^2-a^2=l^2, \tag{1.2}$$

where w is the magnitude of the velocity and k, l are constants. Further, it can be shown that

$$(w/w_1)^2=1+M_1^{-2}[(2/\gamma)+(1/\gamma^2)], \tag{1.3}$$

where M_1 is the Mach number of the flow for $w=w_1$, and $\gamma=1.4$. Finally, it may be shown that the angle, ω, between the two families of characteristics in the physical plane is related to the magnitude of the velocity by

$$w=l\sec\tfrac{1}{2}\omega. \tag{1.4}$$

2. THE PHYSICAL (x, y) PLANE

We shall consider a gas jet which emerges from straight walls and is bounded by two free stream lines. As in the subsonic case,[3] we shall assume that along each free stream line, the magnitude of the velocity is constant. In Fig. 1, AB and $A'B'$ represent the two straight walls from which the jet emerges. Along each of these lines, the stream function ψ is constant. Further, AC and $A'C'$ represent the two free stream lines bounding

[1] F. Frankl, Bull. Acad. Sci. U.R.S.S., Ser. Math. (Izvestia Akad. Nauk S.S.S.R.), 9, 121–143 (1945).

[2] N. Coburn, Quart. Appl. Math. 3, No. 2 (July, 1945).

[3] S. A. Chaplygin, "On gas jets," Sci. Ann. Imp. Univ. Moscow, Physico-Math. Division, Pub. No. 21 (Moscow, 1904), translated from the Russian by M. Slud (Brown University Notes, Providence, Rhode Island, 1944).

the jet. Along each of these free stream lines, the magnitude of the velocity, w, and the stream function, ψ, are constant. From symmetry considerations, we deduce that the velocity vector has the same magnitude on both free stream lines. Further, from symmetry, we see that the y-axis (line ED) is a stream line along which ψ is constant.

The following values will be assigned to the stream constants:

$$B'A'C',\ \psi=M/2;\ ED,\ \psi=0;\ BAC,\ \psi=-M/2. \quad (2.1)$$

If ρ is the density and (u, v) are the x and y components, respectively, of the velocity vector, then

$$M=-\int_{A'}^{A} d\psi=\int_{A'}^{A} \rho v dx-\rho u dy. \quad (2.2)$$

Evidently, M is the mass of gas entering the jet per unit time.

We shall specify the inclination of the walls by use of the negative angle $\tilde{\theta}$ (see Fig. 1). A and A' are situated at the jet entrance, where the velocity vector has the magnitude, w_2. The points B and B' are two symmetrically situated points on the walls where the velocity vector has the magnitude, w_1. For given values of w_2 and θ, the quantity w_1 will be fixed by a relation to be determined (see Eq. 3.9). Further, the curves $B'E$ and EB will represent the locus of points along which $w=w_1$. Evidently, the curves $B'E$ and EB are symmetric. The points C and C', which are symmetrically situated on the two free stream lines, and the point D will be fixed by specifying their hodograph coordinates.

Evidently, we may write Eq. (1.2) in the form

$$l^2=w_1^2-a_1^2=w_2^2-a_2^2, \quad (2.3)$$

where a_1, a_2 are the local sound speeds corresponding to w_1, w_2, respectively. By use of the angle ω, one can simultaneously treat the cases: (1) $w_1<w_2$; (2) $w_1>w_2$. From Eq. (1.4), it follows that these two cases can be characterized by: (1) $0\leqq\omega_1<\omega_2<\pi$; (2) $-\pi<\omega_1<\omega_2\leqq 0$. In mapping from the (ω, θ) plane to the (ϕ, ψ) plane, we shall assume only that $\omega_1<\omega_2$. Hence, this phase of our work will be applicable to either of the above cases. However, it will be shown that in order for the (ϕ, ψ) plane to map into the proper region of the (x, y) plane, the angle ω must be negative ($w_2<w_1$). Finally, we note that by use of Eq. (1.3) an upper bound for the ratio w/w_2 can be given in terms of the Mach number at A or A'.

3. THE HODOGRAPH (ω, θ) PLANE

We introduce the variable θ which denotes the angle between the velocity vector and the x-axis at any point of the physical (x, y) plane. The plane determined by ω and θ will be called the hodograph plane. From Fig. 1, the hodograph map can be determined. We see that

$$\begin{aligned} &B'A': \quad \theta=\tilde{\theta};\\ &ED: \quad \theta=-\pi/2;\\ &AB: \quad \theta=-\pi-\tilde{\theta}. \end{aligned} \quad (3.1)$$

Further, from the discussion in the preceding section, we find that

$$A, A': \quad \omega=\omega_2; \quad B'EB: \quad \omega=\omega_1. \quad (3.2)$$

In order properly to specify the coordinates of D, C, C', it will be necessary to discuss some of the properties of the stream function and the velocity potential in the (ω, θ) plane. We shall prove:

Theorem 1. If the mapping functions connecting the (ϕ, ψ) and the (ω, θ) planes are continuous, with second partial derivatives and nonvanishing jacobian, then ψ and ϕ satisfy the wave equation.

From our previous work,[2] we find

$$\omega=F\left(\frac{\phi}{2l}+\frac{\psi}{2k}\right)+G\left(\frac{\phi}{2l}-\frac{\psi}{2k}\right), \quad (3.3)$$

$$2\theta=F\left(\frac{\phi}{2l}+\frac{\psi}{2k}\right)-G\left(\frac{\phi}{2l}-\frac{\psi}{2k}\right), \quad (3.4)$$

where F and G are arbitrary functions of their respective arguments which possess second partial derivatives (at least sectionally), l is the constant defined by Eq. (2.3) and k is the constant, $\rho a=\rho_1 a_1=\rho_2 a_2$. If the jacobian of Eqs. (3.3), (3.4) does not vanish, we may solve these equations for ϕ, ψ. We find

$$\frac{\phi}{l}=f\left(\frac{\omega}{2}+\theta\right)+g\left(\frac{\omega}{2}-\theta\right), \quad (3.5)$$

$$\frac{\psi}{k}=f\left(\frac{\omega}{2}+\theta\right)-g\left(\frac{\omega}{2}-\theta\right), \quad (3.6)$$

where f and g are the inverse functions to F and G, respectively. From Eq. (3.6), we see that the stream function (as well as the velocity potential) is a solution of the wave equation

$$(\partial^2\psi/\partial\theta^2)-4(\partial^2\psi/\partial\omega^2)=0. \quad (3.7)$$

We shall now determine the hodograph coordinates of D, C, C'. The slopes of the characteristic lines of

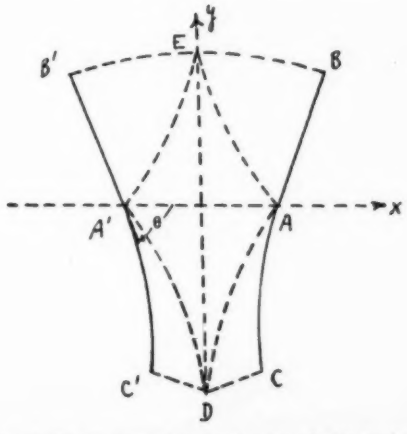


FIG. 1. Flow through a jet in the physical plane.

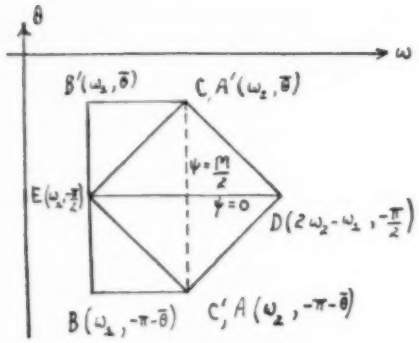


FIG. 2. The hodograph map of the flow.

Eq. (3.7) are

$$d\theta/d\omega = \pm\tfrac{1}{2}. \tag{3.8}$$

First, we require that the point A' lie on the characteristic through E. Hence, the variables $\bar{\theta}$, ω_1, ω_2 satisfy the equation

$$\bar{\theta}+\pi/2=(\omega_2-\omega_1)/2. \tag{3.9}$$

The point D will be fixed as the intersection of the characteristic of negative slope passing through A' and the line $\theta=-\pi/2$. The resulting coordinates of D are indicated on Fig. 2. If $2\omega_2-\omega_1>\pi$ for positive ω (or $2\omega_2-\omega_1>0$ for negative ω), then the point D must be replaced by two points D and D'. These latter points would be the intersections of characteristics through A and A', respectively, with the line $\omega=\pi$ for positive ω (or $\omega=0$ for negative ω). For simplicity, we assume $2\omega_2-\omega_1<\pi$ for positive ω (or $2\omega_2-\omega_1<0$ for negative ω). C and C' will be fixed by assuming that these points coincide with A' and A, respectively.

In the following sections, we shall consider only three regions of the (ω, θ) plane: $EB'A'$ (henceforth denoted as region I); $EA'D$ (henceforth denoted as region II); $A'C'D$ (henceforth denoted as region III). The regions EBA, EAD, ACD may be treated by symmetry considerations. Further, the extension of our method to other regions will be apparent.

Next, we consider the boundary conditions on ψ for the integration of Eq. (3.7). In addition to the boundary conditions (2.1), data along $B'E$ must be given. We shall assume boundary conditions along $B'E$ which are equivalent to Cauchy data

$$B'E:\ \psi=\tfrac{1}{2}M+kh(\theta),\ \phi=lp(\theta),\ -\pi/2\le\theta\le\bar{\theta}, \tag{3.10}$$

where k and l are the constants previously introduced and $h(\theta)$ and $p(\theta)$ possess second derivatives. In order for the stream function to be continuous along $B'E$, it is necessary that

$$h(\bar{\theta})=0,\quad h(-\pi/2)=-M/2k. \tag{3.11}$$

We shall now show that:

Theorem 2. A solution for ψ exists in region I with the following properties: (1) the solution depends upon a single function f of the variables $\frac{1}{2}(\omega-\omega_2)+\theta-\bar{\theta}$ and $\frac{1}{2}(\omega-\omega_2)-\theta+\bar{\theta}$; (2) the range of θ in $f(\theta)$ is $\omega_1-\omega_2\le\theta\le 0$.

Replacing ψ by $\psi-M/2$, ω by $\omega-\omega_2$, θ by $\theta-\bar{\theta}$ in Eqs. (3.5), (3.6), we obtain

$$\psi=\frac{M}{2}+k\left[f\left(\frac{\omega-\omega_2}{2}+\theta-\bar{\theta}\right)-g\left(\frac{\omega-\omega_2}{2}-\theta+\bar{\theta}\right)\right], \tag{3.12}$$

$$\phi=l\left[f\left(\frac{\omega-\omega_2}{2}+\theta-\bar{\theta}\right)+g\left(\frac{\omega-\omega_2}{2}-\theta+\bar{\theta}\right)\right]. \tag{3.13}$$

In order to determine the range of the arguments of f and g, we write

$$\tfrac{1}{2}(\omega-\omega_2)+\theta-\bar{\theta}=\theta_0,\quad \tfrac{1}{2}(\omega-\omega_2)-\theta+\bar{\theta}=\theta_1. \tag{3.14}$$

It is easily shown that the characteristic lines (3.14) will intersect region I if and only if

$$\omega_1-\omega_2\le\theta_0\le\theta_0,\quad \tfrac{1}{2}(\omega_1-\omega_2)\le\theta_1\le 0. \tag{3.15}$$

Thus, the arguments of f and g have the common range $\frac{1}{2}(\omega_1-\omega_2)$ to 0. In this common range, it follows from Eq. (3.12) and the boundary condition, $\psi=\frac{1}{2}M$ when $\theta=\bar{\theta}$, that $f(\theta)=g(\theta)$ for $\frac{1}{2}(\omega_1-\omega_2)\le\theta\le 0$. Thus, we may consider f and g as single function, f, defined over the interval $\omega_1-\omega_2$ to 0.

The boundary conditions (3.10) furnish relations for determining $f(\theta)$. Thus, from Eqs. (3.10), (3.12), (3.13), we find

$$2f\left(\frac{\omega_1-\omega_2}{2}+\theta\right)=h(\theta+\bar{\theta})+p(\theta+\bar{\theta}), \tag{3.16}$$

$$2f\left(\frac{\omega_1-\omega_2}{2}-\theta\right)=-h(\theta+\bar{\theta})+p(\theta+\bar{\theta}). \tag{3.17}$$

Equation (3.16) determines $f(\theta)$ for

$$\omega_1-\omega_2\le\theta\le\tfrac{1}{2}(\omega_1-\omega_2);$$

Eq. (3.17) determines $f(\theta)$ for $\frac{1}{2}(\omega_1-\omega_2)\le\theta\le 0$. Further, the solutions for ψ and ϕ in region I may be written in terms of this function, f, as

$$\text{I:}\quad \psi=\frac{M}{2}+k\left[f\left(\frac{\omega-\omega_2}{2}+\theta-\bar{\theta}\right)-f\left(\frac{\omega-\omega_2}{2}-\theta+\bar{\theta}\right)\right], \tag{3.18}$$

$$\text{I:}\quad \phi=l\left[f\left(\frac{\omega-\omega_2}{2}+\theta-\bar{\theta}\right)+f\left(\frac{\omega-\omega_2}{2}-\theta+\bar{\theta}\right)\right]. \tag{3.19}$$

It should be noted that the function $f(\theta)$ as determined by Eq. (3.16) for the range $\omega_1-\omega_2\le\theta\le\frac{1}{2}(\omega_1-\omega_2)$

may differ completely in analytic form from the function $f(\theta)$ as determined by Eq. (3.17) for the range $\frac{1}{2}(\omega_1-\omega_2)\leq\theta\leq 0$. However, by use of Eqs. (3.16), (3.17), it can be easily shown that:

Theorem 3. The necessary and sufficient conditions that $f(\theta)$ be represented by the same function in the above two intervals are that $h(\theta)$ have odd symmetry about $\theta=\bar{\theta}$ and $p(\theta)$ have even symmetry about $\theta=\bar{\theta}$.

We shall continue our solution for ψ into regions II ($EA'D$) and III ($A'C'D$) by use of the well-known methods[4] for integrating the wave equation. The solution for ϕ in these regions can then be determined by use of Eq. (3.5) and continuity considerations. We find

$$\text{II:}\quad \psi=k\left[f\left(\frac{\omega-\omega_2}{2}+\theta-\bar{\theta}\right)-f\left(\frac{\omega-\omega_2-2\pi}{2}-\theta-\bar{\theta}\right)\right], \tag{3.20}$$

$$\text{II:}\quad \phi=\frac{Ml}{2k}+l\left[f\left(\frac{\omega-\omega_2}{2}+\theta-\bar{\theta}\right)+f\left(\frac{\omega-\omega_2-2\pi}{2}-\theta-\bar{\theta}\right)\right], \tag{3.21}$$

$$\text{III:}\quad \psi=\frac{M}{2}+k\left[f\left(\frac{-\omega+\omega_2-2\pi}{2}-\theta+\bar{\theta}\right)-f\left(\frac{\omega-\omega_2-2\pi}{2}-\theta-\bar{\theta}\right)\right], \tag{3.22}$$

$$\text{III:}\quad \phi=\frac{Ml}{k}+l\left[f\left(\frac{-\omega+\omega_2-2\pi}{2}-\theta+\bar{\theta}\right)+f\left(\frac{\omega-\omega_2-2\pi}{2}-\theta-\bar{\theta}\right)\right]. \tag{3.23}$$

It is easily shown that the functions in the right-hand side of the above equations are defined by Eqs. (3.16), (3.17) in regions II and III.

With the aid of Eqs. (3.20) through (3.23), we shall prove three results relating the (ω, θ) and (ϕ, ψ) plane maps of the flow.

Theorem 4. The necessary and sufficient condition for the continuity across EA' of the mapping functions ψ and ϕ is that $h(-\pi/2)=-M/2k$; the necessary and sufficient condition that the first partial derivatives of ψ and ϕ be continuous across EA' is that $p'(-\pi/2)=0$.

To show the necessity of the condition $h(-\pi/2)=-M/2k$, we compute the values of Eqs. (3.18), (3.19), (3.20), (3.21) along EA'. We find

$$\psi(\text{II})-\psi(\text{I})=-\frac{M}{2}+k\left[f\left(\frac{\omega_1-\omega_2+\pi}{2}+\bar{\theta}\right)-f\left(\frac{\omega_1-\omega_2-\pi}{2}-\bar{\theta}\right)\right], \tag{3.24}$$

$$\phi(\text{II})-\phi(\text{I})=\frac{Ml}{2k}-l\left[f\left(\frac{\omega_1-\omega_2+\pi}{2}+\bar{\theta}\right)-f\left(\frac{\omega_1-\omega_2-\pi}{2}-\bar{\theta}\right)\right]. \tag{3.25}$$

By use of Eqs. (3.9), (3.16), (3.17), we obtain

$$f\left(\frac{\omega_1-\omega_2+\pi}{2}+\bar{\theta}\right)-f\left(\frac{\omega_1-\omega_2-\pi}{2}-\bar{\theta}\right)=f(0)-f(\omega_1-\omega_2)=-h\left(\frac{-\pi}{2}\right). \tag{3.26}$$

Comparing the last three equations, we see the necessity of our condition. The sufficiency of this condition follows from the fact that a continuous solution has been constructed with the aid of this condition.

In order to verify the second part of our result, we compute the derivatives of ψ and ϕ in regions I and II. Denoting the jump of the derivative across EA' by brackets, and using Eqs. (3.9), (3.16), (3.17), we find

$$f'(\omega_1-\omega_2)-f'(0)=p'\left(-\frac{\pi}{2}\right)=\frac{1}{k}\left[\frac{\partial\psi}{\partial\theta}\right]=-\frac{2}{k}\left[\frac{\partial\psi}{\partial\omega}\right]=-\frac{1}{l}\left[\frac{\partial\phi}{\partial\theta}\right]=\frac{2}{l}\left[\frac{\partial\phi}{\partial\omega}\right]. \tag{3.27}$$

The primes on f and p denote differentiation of these functions with respect to their respective arguments. Equation (3.27) verifies our second result.

In the same manner, we can prove:

Theorem 5. The necessary and sufficient condition for the continuity of ψ and ϕ across $A'D$ (the boundary of II and III) is that $h(-\pi/2)=-M/2k$; the necessary and sufficient condition for the continuity across $A'D$ of the first partial derivatives of ψ and ϕ is that $h'(-\pi/2)=0$.

Finally, we examine the Jacobian of the transformation from the (ω, θ) to the (ϕ, ψ) plane. We shall prove:

Theorem 6. The necessary and sufficient conditions that the jacobian, $J(\phi, \psi/\omega, \theta)$, shall not vanish in the closed regions I, II, III are $h'(\theta)+p'(\theta)\neq 0$, $h'(\theta)-p'(\theta)\neq 0$.

[4] J. D. Tamarkin and W. Feller, *Partial Differential Equations* (Brown University Notes, Providence, Rhode Island, 1941), pp. 24-28.

By use of the expressions for ψ and ϕ in region I, we find

$$J\left(\frac{\phi,\psi}{\omega,\theta}\right) = 2klf'\left(\frac{\omega-\omega_2}{2}+\theta-\tilde{\theta}\right)f'\left(\frac{\omega-\omega_2}{2}-\theta+\tilde{\theta}\right). \quad (3.28)$$

Similar expressions are valid for regions II and III. Thus, the nonvanishing of the jacobian is equivalent to the nonvanishing of $f'(\theta)$. From Eqs. (3.16), (3.17), it follows that this latter condition is satisfied if and only if $h'(\theta)\pm p'(\theta)\neq 0$.

4. SOME MAPPING RELATIONS

The functions relating the (ϕ, ψ) and the (x, y) plane can be analyzed by use of Eq. (1.4). If u and v are the

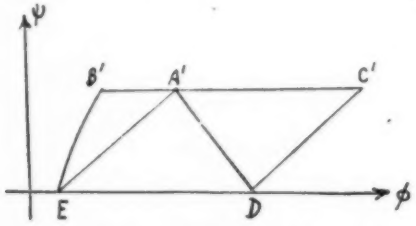


FIG. 3. The $\phi-\psi$-plane map of the flow.

x and y components of the velocity, respectively, and ρ is the density, then[2]

$$u=\frac{\partial\phi}{\partial x}=-\frac{1}{\rho}\frac{\partial\psi}{\partial y}=l\sec\tfrac{1}{2}\omega\cos\theta, \quad (4.1)$$

$$v=\frac{\partial\phi}{\partial y}=-\frac{1}{\rho}\frac{\partial\psi}{\partial x}=l\sec\tfrac{1}{2}\omega\sin\theta. \quad (4.2)$$

From the known relations $a=w\sin\omega/2$, $\rho a=k$, where w is the magnitude of the velocity vector, one can determine ρ as a function of ω and the above relations may be written as

$$\frac{\partial y}{\partial\psi}=-\frac{1}{k}\sin\tfrac{1}{2}\omega\cos\theta, \quad \frac{\partial x}{\partial\psi}=-\frac{1}{k}\sin\tfrac{1}{2}\omega\sin\theta, \quad (4.3)$$

$$\frac{\partial y}{\partial\phi}=\frac{1}{l}\cos\tfrac{1}{2}\omega\sin\theta, \quad \frac{\partial x}{\partial\phi}=\frac{1}{l}\cos\tfrac{1}{2}\omega\cos\theta. \quad (4.4)$$

In the next few paragraphs, we shall be concerned with the (ϕ, ψ) plane map of the flow. Evidently, the line ED and the curve $B'A'C'$ (see Fig. 1) map into the lines $\psi=0$ and $\psi=M/2$ in the (ϕ, ψ) plane. We shall show:

Theorem 7. If we move from B' to A', A' to C', E to D (see Fig. 1) then ϕ is increasing. That is, the maps of the boundaries $B'A'C'$ and ED in the (x, y) and (ϕ, ψ) planes are in one-to-one correspondence.

To verify this result, we note that in moving from B' to A' along the wall, x increases and y decreases. From (4.4) and the fact that $\theta<0$ along $B'A'$, it follows that $\partial x/\partial\phi>0$, $\partial y/\partial\phi<0$. Hence, ϕ must increase as one moves from B' to A'. The other results may be shown in similar manner.

The following result is an immediate consequence of the assumptions and the form of the equations for ϕ and ψ in the various regions (or the formulas 3.3 through 3.6).

Theorem 8. If the boundaries $B'A'C'$ and ED of the (ω, θ) and (ϕ, ψ) planes are in one-to-one correspondence and if the jacobian, $J(\phi, \psi/\omega, \theta)$, does not vanish over the regions I, II, III, then: (1) the characteristics of the (ω, θ) plane transform into two families of straight lines (which will be called characteristics) in the (ϕ, ψ) plane; (2) to each characteristic of positive (negative) slope in (ω, θ) plane, there corresponds a unique characteristic of positive (negative slope) in (ϕ, ψ) plane and conversely.

If the above assumptions are valid, then by use of our results, one may immediately construct the (ϕ, ψ) plane map of the flow (see Fig. 3). In order to obtain conditions under which the boundaries $B'A'C'$ and ED of the (ω, θ) and (ϕ, ψ) planes are in one-to-one correspondence, we shall prove:

Theorem 9. The necessary and sufficient conditions that the boundaries $B'A'C'$ and ED of the (ω, θ) and (ϕ, ψ) planes be in one-to-one correspondence are $h'(\theta)+p'(\theta)>0$, $h'(\theta)-p'(\theta)>0$.

Consider the line $B'A'$. In moving from B' to A', we have shown in theorem 7 that ϕ is increasing. From Eq. (3.19), we find that along $B'A'$

$$\phi=2lf\left(\frac{\omega-\omega_2}{2}\right), \quad \omega_1\leq\omega\leq\omega_2. \quad (4.5)$$

Equation (4.5) implies that the range of θ in $f(\theta)$ is $\frac{1}{2}(\omega_1-\omega_2)\leq\theta\leq 0$. Hence, the $f(\theta)$ as defined by Eq. (3.17) must be used. Replacing θ in Eq. (3.17) by $\frac{1}{2}(\omega_1-\omega_2)$, we obtain

$$\phi=2lf\left(\frac{\omega-\omega_2}{2}\right) = 2l\left[-h\left(\tilde{\theta}+\frac{\omega_1+\omega}{2}\right)+p\left(\tilde{\theta}+\frac{\omega_1-\omega}{2}\right)\right]. \quad (4.6)$$

Since in moving from B' to A', both ϕ and ω are increasing (ω is increasing, since $-\pi<\omega_1<\omega_2<0$, or $0<\omega_1<\omega_2<\pi$), then $d\phi/d\omega>0$. By differentiation of Eq. (4.6), we find that this last statement is equivalent to

$$h'(\theta)-p'(\theta)>0, \quad -\pi/2\leq\theta\leq\tilde{\theta}. \quad (4.7)$$

A similar argument for the boundaries $A'C'$ and ED will verify the necessity of the remaining condition.

Further, by reversing the order of the above steps, one may verify the sufficiency of the above conditions.

From theorem 9, it follows that:

Theorem 10. If $h'(\theta)>0$, $p'(\theta)>0$ (or $p'(\theta)<0$) and $h'(\theta)>p'(\theta)$, then the (ω, θ) and (ϕ, ψ) maps of the boundaries of regions I, II, III are in one-to-one correspondence.

The proof can be given in three steps. First, the above conditions imply $h'(\theta)\pm p'(\theta)>0$ and hence that $f(\theta)$ is an increasing function of θ [see Eqs. (3.16, 3.17)]. Thus, by theorem 9, the maps of ED, $B'A'C'$, are in one-to-one correspondence. From Eq. (3.10) and our conditions, we see that the maps of EB' are in one-to-one correspondence. Finally, the mapping functions along the characteristic EA', which bounds region I, have the values

$$EA': \quad \begin{aligned} \psi &= \tfrac{1}{2}M+k[f(\omega-\omega_2)-f(0)], \\ \phi &= l[f(\omega-\omega_2)+f(0)]. \end{aligned} \tag{4.8}$$

Since $f(\theta)$ is an increasing function of θ, the maps of EA' are in one-to-one correspondence. In a similar manner, it may be shown that the (ω, θ) and (ϕ, ψ) maps of $A'D$, $C'D$ are in one-to-one correspondence.

Finally, we consider the mapping of the interior points of the (ω, θ) and (ϕ, ψ) planes. We prove:

Theorem 11. If the boundaries of the (ω, θ) and (ϕ, ψ) planes are in one-to-one correspondence (see theorem 10), then the interior points of these two maps of the flow are in one-to-one correspondence.

Consider a point P in the interior of region I in the (ω, θ) plane. Through P one can pass two characteristics of the (ω, θ) plane. These characteristics will be such that they will intersect either $B'A'$ or EB' in two points. Consider the case where the lines intersect EB' in the points Q and R. Using subscripts to denote the values of the variables at various points, we see that $\theta_Q<\theta_P<\theta_R$. Since Q and R lie on characteristics through P, we find that (ω_P, θ_P) satisfy

$$\begin{aligned} (\theta_P-\theta_R)/(\omega_P-\omega_R) &= -\tfrac{1}{2}, \\ (\theta_P-\theta_Q)/(\omega_P-\omega_Q) &= \tfrac{1}{2}. \end{aligned} \tag{4.9}$$

Substituting Eq. (4.9) into the formula for ψ in region I and noting that $\omega_R=\omega_Q=\omega_1$, we find

$$\psi_P=\frac{M}{2}+k\left[f\left(\frac{\omega_1-\omega_2}{2}+\theta_R-\tilde{\theta}\right)-f\left(\frac{\omega_1-\omega_2}{2}+2\theta_R-2\theta_P-\theta_R+\tilde{\theta}\right)\right], \tag{4.10}$$

$$\psi_P=\frac{M}{2}+k\left[f\left(\frac{\omega_1-\omega_2}{2}+2\theta_P-2\theta_Q+\theta_Q-\tilde{\theta}\right)-f\left(\frac{\omega_1-\omega_2}{2}-\theta_Q+\tilde{\theta}\right)\right], \tag{4.11}$$

$$\psi_R=\frac{M}{2}+k\left[f\left(\frac{\omega_1-\omega_2}{2}+\theta_R-\tilde{\theta}\right)-f\left(\frac{\omega_1-\omega_2}{2}-\theta_R+\tilde{\theta}\right)\right], \tag{4.12}$$

$$\psi_Q=\frac{M}{2}+k\left[f\left(\frac{\omega_1-\omega_2}{2}+\theta_Q-\tilde{\theta}\right)-f\left(\frac{\omega_1-\omega_2}{2}-\theta_Q+\tilde{\theta}\right)\right]. \tag{4.13}$$

By noting that $f(\theta)$ is an increasing function of θ and $\theta_P-\theta_Q>0$, $\theta_R-\theta_P>0$, we see from Eqs. (4.10) through (4.13) that $\psi_Q<\psi_P<\psi_R$. A similar argument shows that $\phi_P>\phi_Q$, $\phi_P>\phi_R$, and also that P must lie to the left of the intersection of the characteristics RP and EA'. Further, the same argument will show that interior points of regions II and III of the (ω, θ) plane map into interior points of the corresponding regions of the (ϕ, ψ) plane.

Again, since the inverse function to $f(\theta)$ is also an increasing function, it can be shown that a point P in the (ϕ, ψ) plane interior maps into a unique point P in the (ω, θ) plane interior.

Our final results are concerned with the mapping from the (x, y) plane to the (ϕ, ψ) plane. We shall show that it is essential that ω be negative (or $w_2<w_1$). In particular, we shall prove:

Theorem 12. If, (1) no back flow exists (θ is negative), (2) $w_2<w_1$ (ω is negative), (3) the (x, y) plane map of EB' is in one-to-one correspondence with (ϕ, ψ) plane map of EB', (4) $h'(\theta)>0$, $p'(\theta)>0$ (or $p'(\theta)<0$) $h'(\theta)\pm p'(\theta)>0$, then there exists a one-to-one correspondence between the interior points of the (x, y) and (ϕ, ψ) planes.

Through any point P in the interior of the (ϕ, ψ) plane, there pass two lines, the line $\phi=$constant and the line $\psi=$constant. These lines intersect the boundaries, $B'E$ and $B'A'C'$, in the points, R and S, respectively. By theorem 7 and assumption (3), the boundaries of the (x, y) and (ϕ, ψ) planes are in one-to-one correspondence. Hence, the points R and S have unique maps in the (x, y) plane. Evidently, $\psi_P<\psi_S$, $\phi_P>\phi_R$. From assumptions (1) and (2) of the theorem and Eqs. (4.3), (4.4), it is seen that along SP, $\partial x/\partial\psi<0$, and along RP, $\partial y/\partial\phi<0$. Thus, $x_P>x_S$, $y_P>y_R$, and to each interior point of the (ϕ, ψ) plane, there corresponds at least one interior point of the (x, y) plane.

Further, it is easily seen that a unique interior point corresponding to P must exist. From Eqs. (4.3), (4.4), we find

$$x_P-x_S=\int_{\psi_S}^{\psi_P}-\frac{1}{k}\sin\frac{\omega}{2}\sin\theta\,d\psi, \tag{4.14}$$

$$y_P-y_R=\int_{\phi_R}^{\phi_P}\frac{1}{l}\cos\frac{\omega}{2}\sin\theta\,d\phi. \tag{4.15}$$

By assumption (4) and theorems 10 and 11, to each interior point of the (ϕ, ψ) plane, there corresponds a unique interior point of the (ω, θ) plane. Hence, the above integrals are single-valued and define unique values of (x_P, y_P). A similar argument shows that to each interior point in the (x, y) plane, there corresponds a unique interior point of the (ϕ, ψ) plane.

5. THE FUNCTIONS CONNECTING THE (ω, θ) AND (x, y) PLANES

We shall merely state some results. By use of Eqs. (4.3), (4.4), the expressions (3.18) through (3.23), and the chain rule for differentiation, we can easily obtain the expressions for $\partial x/\partial\omega$, $\partial x/\partial\theta$, $\partial y/\partial\omega$, $\partial y/\partial\theta$ in the regions I, II, III. It can be shown with the aid of some detailed computation that:

Theorem 13. The partial derivatives $\partial x/\partial\omega$, $\partial x/\partial\theta$, $\partial y/\partial\omega$, $\partial y/\partial\theta$ are such that; (1) the integrability conditions are satisfied; (2) the tangential derivatives of the function $x(\omega, \theta)$ along EA' in region I is equal to the tangential derivative of this function along EA' in region II (similar results are valid for $y(\omega, \theta)$ along EA' and for both functions along $A'D$); (3) thus, these functions define a continuous map from the (ω, θ) plane to the (x, y) plane.

6. SOME REMARKS

It should be noted that our methods can be extended to other regions of the (x, y) plane. The fundamental idea can be obtained by use of Fig. 3. Evidently, the extension of ED and the two characteristics through C' furnish region IV. Thus, region IV in the (ω, θ) plane should be the region bounded by ED, EC', and $C'D$. This folding property of the (ω, θ) plane map will furnish a type of periodic recurrence of the jet boundary in the physical (x, y) plane.

Finally, it should be noted that the above mapping idea is applicable to the two-dimensional jet for an arbitrary gas law. Here, the problem is to determine the Riemann function in the hodograph plane.[5]

[5] S. Bergman, Trans. Am. Math. Soc. **57**, No. 3, 299–331 (May, 1945).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

A Mechanical Determination of Biaxial Residual Stress in Sheet Materials

R. G. TREUTING AND W. T. READ, JR.
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received June 5, 1950)

A method is given for determining the residual stress in a sheet material by removing successive uniform layers of material from the surface of a test specimen and measuring the resulting curvature. From the condition of equilibrium of a free specimen, a stress *vs* curvature relation is derived which holds over the depth to which material has been removed. The method applies when the stress is constant in the plane of the specimen and varies through the thickness. An experimental technique is described which is believed to satisfy the essential requirement that the removal of surface layers should not affect the stress in the remaining material, and a practical example is given.

I. INTRODUCTION

RESIDUAL stresses, that is, stresses which may exist in a material free of external load, generally result from nonhomogeneous plastic deformation, which may be introduced in many ways including probably all of the fabrication methods.[1] These stresses have an important effect on the physical properties of metals. Fatigue life in bending, for example, depends on the maximum stress developed at the surface including residual stress. The important practical problems of measuring residual stresses have been attacked in many ways. One method, for which a bibliography is given,[1] involves unbalancing the self-equilibrium of the internal stresses by removing or sectioning a part of the material and observing the resulting deformation.[2–4] The present treatment of this method differs from the previous ones in two respects: first, two principal components of stress are considered; second, these components are explicitly expressed as functions of curvature, thereby eliminating the need for calculation by successive approximations.

The method applies to sheet materials and involves removing successive uniform layers from the surface of a sample. In an initially straight sample symmetry makes it necessary to remove material to a depth of only half the thickness. When the following three conditions are satisfied, the accuracy of the method is limited only by the precision of the measurements.

(i) In order to determine the stress from the curvature it is necessary that the sample be linear in pure bending, over the range of curvatures involved, and that the elastic constants be the same throughout the material. If these conditions are not satisfied, the stress cannot be obtained from the curvature. Instead, it is necessary to measure the bending moment required to straighten the strip after every removal of material. A

[1] W. M. Baldwin, Jr., "Residual stresses in metals," Edgar Marburg Lecture, A.S.T.M., 1949.
[2] N. Davidenkov, Z. Metallkunde **24**, 25 (1932).
[3] G. Sachs and G. Espey, Trans. AIME **147**, 348 (1942).
[4] D. G. Richards, Proc. Soc. Exptl. Stress Analysis **3**, No. 1, 40 (1945).

stress-moment relation which is independent of the properties of the material is given in the Appendix.

(ii) The stress should not vary in the plane of the specimen but only through the thickness. The method is therefore not suitable for detecting critical stress concentrations in localized areas which are smaller than the area of the strip.

(iii) The removal of surface layers should not disturb the stress in the remaining material. A technique which is believed to satisfy this condition is described in Sec. IV.

Although the derivation given in Sec. II assumes that the material is isotropic, the final formulas for the general, anisotropic, case are given in the Appendix.

Condition (ii) above is equivalent to the condition that the components of stress normal to the plane of the specimen shall vanish. Hence, the stress is defined by the directions and magnitudes of the principal components of stress in the plane of the specimen. The principal stress axes coincide with the principal axes of curvature.

The two principal components of curvature are measured and plotted as functions of the depth of material removed. In Sec. II formulas are derived relating stress and curvature over the depth to which material has been removed. These formulas involve the derivative and integral of the curvature, which are measured graphically from the experimental curves.

The formulas are derived from the fact that, in a free specimen, the internal stresses are self-equilibrating; that is, there are no resultant forces or moments on any cross section taken through the specimen. Removing a part of the stress distribution, as in removing a layer of material, leaves an unbalanced distribution of internal stress so that external restraints are required to prevent deformation. When the restraints are released, the specimen so deforms as to restore internal equilibrium, the forces and moments associated with the deformation being equal and opposite to those of the residual stresses.

For example, consider a cold-rolled metal sheet so deformed by compression on its surfaces as to produce a large permanent elongation. In rolling, the contact surfaces are constrained by friction to move at more or less the constant peripheral velocity of the rolls, while the central fibers are compressed beyond the yield strength and flow plastically in the rolling direction. When the sheet leaves the rolls, the elongation is essentially uniform. The difference in the plastic deformation of the outer and central fibers results in the former being held in a residual tensile stress by the latter, which are accordingly compressed. If now a uniform thin layer of material is removed from one surface of a sample strip cut (e.g., longitudinally) from such a sheet, there results a net compressive internal force in the rolling direction, and the strip assumes an elongation to introduce a tensile force equal to that removed. Also, removal of the surface tensile force produces an unbalanced internal bending moment, so that the strip bends in order to restore rotational equilibrium by introducing added tension on the side of removal. This results in curvature in the rolling direction with this side convex.

II. THE RELATION OF RESIDUAL STRESS TO CURVATURE

In this section we shall derive a formula for the stress distribution throughout the depth to which material has been removed. Figure 1 shows a rectangular section of the specimen and a coordinate system with the xy plane in the initial midsurface of the specimen, and the x and y axes in the directions of principal stress. The initial upper and lower surfaces of the specimen are the planes $z=z_0$ and $z=-z_0$. The variable upper surface is the plane $z=z_1$.

Since we are considering the case in which the stress distribution is constant in the x and y directions, the only nonvanishing stress components are σ_x and σ_y (Fig. 1). The equilibrium conditions[5] are satisfied independently of the variation of $\sigma_x=\sigma_x(z)$ and $\sigma_y=\sigma_y(z)$, with z.

Actual test specimens are cut from larger sheets, and except for a negligibly small region near the edges, where the stress vanishes, the stress distribution in the sample is the same as that in the original sheet.

Since the external force and moment on the free specimen vanish, equilibrium requires that the force and moment acting on any internal cross section shall vanish. For a cross section normal to the x axis, Fig. 1, on which the only stress acting is σ_x, there results

$$\int_{-z_0}^{z_0}\sigma_x(z)dz=\int_{-z_0}^{z_0}\sigma_x(z)zdz=0, \qquad (1)$$

and a similar expression in σ_y for the appropriate section.

A uniformly thick layer of material is now removed from the upper surface (shaded area, Fig. 2) so that the new upper surface is at $z=z_1$. Let the specimen be prevented from deforming by the application of external loads to the edges. If the process of removing material introduces no further residual stress, the internal stress distribution remains constant, but the force and mo-

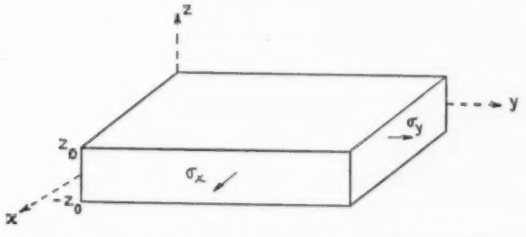


FIG. 1. Element of test specimen showing principal components of residual stress.

[5] Stephen Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1934), p. 195. (Here the body forces X, Y, and Z are negligible.)

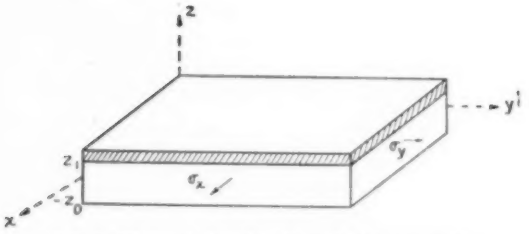


FIG. 2. Element showing layer to be removed.

ment acting on a cross section no longer vanish. Instead, on the cross section normal to the x axis there acts the resultant force per unit width F_x, Fig. 3, and moment per unit width M_x, each a function of the depth of material removed: $F_x=F_x(z_1)$ and $M_x=M_x(z_1)$. Since the only stress on the section remains σ_x (Fig. 1),

$$F_x(z_1)=\int_{-z_0}^{z_1}\sigma_x(z)dz. \tag{2}$$

Remembering that the center line of the cross section has been displaced in the negative z-direction by the amount $(z_0-z_1)/2$, Fig. 4, we have for the moment per unit width about the center line

$$M_x(z_1)=\int_{-z_0}^{z_1}\sigma_x(z)\left[z+\frac{z_0-z_1}{2}\right]dz, \tag{3}$$

with similar expressions for the force and moment on a cross section normal to the y-axis. From Eq. (1), $F_x(z_0)=M_x(z_0)=0$.

The forces F_x and F_y and moments M_x and M_y are equal to the external forces and moments applied to the edges of the specimen to prevent deformation. When these external loads are released, the deformation must be such as to introduce additional stresses which restore internal equilibrium. The forces due to the additional stress introduced by deformation are given by the elongation, and the moments by the curvature. Since the components of curvature in the x and y directions $\varphi_x=\varphi_x(z_1)$ and $\varphi_y=\varphi_y(z_1)$ are measured more accurately as functions of z_1 than are the corresponding elongations, we make use of the moment equation (3) and express $M_x(z_1)$ in terms of the curvature. In the present case of vanishing stress components normal to the plane of the specimen, the formula for pure bending of a flat plate applies exactly, and we have for an isotropic material in the elastic range[6]

$$M_x(z_1)=[-E/12(1-\nu^2)](z_0+z_1)^3[\varphi_x(z_1)+\nu\varphi_y(z_1)], \tag{4}$$

where E is Young's modulus, ν is Poisson's ratio, and the negative sign is used, since the curvature occurs on removal of the applied moment. Substituting Eq. (4) for $M_x(z_1)$ into Eq. (3) gives the desired result, an equation relating the unknown stress σ_x and the measured curvature; and the derivation is complete so far as physical principles are concerned. However, in this form σ_x occurs under the integral sign. To transform the integral equation to an explicit expression for σ_x some mathematical manipulation is required and this is performed in the Appendix. The result is

$$\sigma_x(z_1)=\frac{-E}{6(1-\nu^2)}\left\{(z_0+z_1)^2\left[\frac{d\varphi_x(z_1)}{dz_1}+\nu\frac{d\varphi_y(z_1)}{dz_1}\right]+4(z_0+z_1)[\varphi_x(z_1)+\nu\varphi_y(z_1)]-2\int_{z_1}^{z_0}[\varphi_x(z)+\nu\varphi_y(z)]dz\right\}, \tag{5}$$

from which the residual stress σ_x may be calculated directly from the experimental curves, the derivatives and integrals of the curvature being obtained graphically.

In the derivation the initial curvature has been assumed to vanish: that is $\varphi_x(z_0)=\varphi_y(z_0)=0$; therefore, in practice any initial curvature should be subtracted from subsequent values to obtain φ_x and φ_y.

The expression for φ_y is similarly derived and is the same as Eq. (5) with the subscripts x and y interchanged. Thus, when σ_x has been found, very little additional work is required to find σ_y.

Equation (5) and the corresponding equation for σ_y apply only to an isotropic strip which is linear in pure bending over the range of curvatures resulting from layer removal. If the strip is not linear or if the elastic constants are not the same throughout the thickness, then the linear moment-curvature relation (4) does not apply and the moments $M_x(z_1)$ and $M_y(z_1)$ required to straighten the deformed strip should be measured as functions of z_1 and Eq. (A7) of the Appendix used. Equation (A7) for stress in terms of moment is independent of the properties of the material and applies in the general nonlinear anisotropic case.

For a material which is linear but anisotropic, Eq. (4) is replaced by the formula for pure bending of an anisotropic flat plate, and Eq. (5) is replaced by the more general anisotropic relation (A8) of the Appendix.

Since the elastic constants E and ν, or s_{11}, s_{12}, s_{22} for an anisotropic material, are dependent on history and composition, the best accuracy will be obtained when these parameters are measured on a sample of the material.

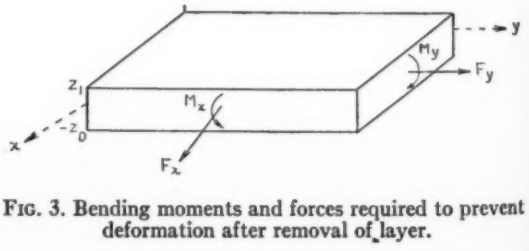


FIG. 3. Bending moments and forces required to prevent deformation after removal of layer.

[6] Stephen Timoshenko, *Theory of Plates and Shells* (McGraw-Hill Book Company, Inc., New York, 1940), p. 41.

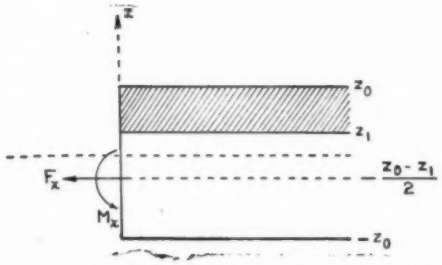


FIG. 4. Cross section showing displacement of midplane.

III. SPECIAL CASES

There are two special cases of practical importance where (5) takes a simpler form and leads to a reduction in the experimental work and calculations. The first case occurs when the strain in one direction can be neglected, as when in rolling there is negligible expansion of the sheet in the transverse (e.g., the y) direction. Then $\sigma_y = \nu\sigma_x$, $\varphi_y = 0$ and Eq. (5) becomes

$$\sigma_x(z_1) = \frac{\sigma_y(z_1)}{\nu} = -\frac{E}{6(1-\nu^2)}\left[(z_0+z_1)^2\frac{d\varphi_x(z_1)}{dz_1} + 4(z_0+z_1)\varphi_x(z_1) - 2\int_{z_1} \varphi_x(z)dz\right], \quad (6)$$

so that only the curvature in the longitudinal direction need be measured.

The second instance when only one component of curvature need be measured occurs when the directional effect in the plane of the specimen can be neglected, as in careful compressional rolling: then $\sigma_x = \sigma_y$, $\varphi_x = \varphi_y$, and Eq. (5) becomes

$$\sigma_x(z_1) = \sigma_y(z_1) = \frac{-E}{6(1-\nu)}\left[(z_0+z_1)^2\frac{d\varphi_x(z_1)}{dz_1} + 4(z_0+z_1)\varphi_x(z_1) - 2\int_{z_1}^{z_0} \varphi_x(z)dz\right]. \quad (7)$$

This type of stress distribution is introduced by shot-peening. For a specimen peened after rolling the situation is more complicated; however, Eq. (7) can still be used to determine the maximum compressive stress, which occurs in the surface fibers, where the peening effect predominates. Below the surface it would be necessary to measure both curvatures and use Eq. (5).

IV. APPLICATION OF THE METHOD

As an example of the method in practice, consider a specimen of "A" phosphor bronze* cold rolled 4 B & S numbers hard (37 percent reduction in thickness) to a final thickness of 0.032 in. The specimen was originally intended for tensile testing and was cut longitudinally to the rolling direction; the 4½-in.×½-in. reduced section was used for the stress analysis.

* Nominal composition is 5 percent Sn, 0.02 to 0.35 percent P, balance copper. It is a single-phase solid solution.

Since the residual stress distribution was considered to result predominantly from the cold-rolling, the simplified Eq. (6) for zero transverse strain was used and curvature measured in the longitudinal (x-) direction only.

The preparation of the specimen consisted of cementing to a toolmaker's steel parallel; grinding on successively finer grades of metallographic abrasive paper supported on plate glass, terminating with No. 600 carborundum; and etching by swabbing with concentrated NH_4OH-5 percent superoxol solution to reduce the grinding marks and reveal the grain structure. With some practice and frequent recourse to a micrometer, the thickness over the significant area can be held uniform to about ±0.0002 in., the majority of the deviation occurring at the corners. The etch employed has a polishing effect and leaves a bright surface. X-ray examination of surfaces thus prepared on recrystallized material shows sharp Debye rings, so that plastic deformation of the surface by the process appears to be minimized.

Cementing is carried out on a small press with steam-heated platens, using a second steel parallel above the specimen, with paper padding between. A clear thermoplastic cement is used; Cenco Sealstix is satisfactory. The surface must be not only free of grease, but in particular of dirt particles: thin spots result on grinding a surface locally elevated. Sufficient pressure is used to flatten the specimen and extrude excess cement; there is an ample margin of safety in this respect. After grinding and etching, the specimen is released on a hot plate and the residual adhesive removed in an appropriate solvent.

It is evident that the temperature to which the specimen is thus subjected must be held below that at which significant stress-relief begins.

Longitudinal curvatures are determined by measuring the sagitta of a chord of known length. The specimen is rested on parallel knife-edge supports a predetermined distance of about half the specimen length apart. The sagitta to this chord is taken as half the difference of height at the midpoint between the convex-up and convex-down positions, to minimize the effect of the specimen's weight. A series of readings are taken at each edge with an optically aligned micrometer height gage. The precision varies among specimens, is generally about ±0.0005 in.

Transverse curvature on narrow specimens can be measured on a metallographic microscope with a calibrated movable stage and calibrated focusing movement, using a large-aperture short-focus lens to reduce the depth of focus.

The specimen in question was thus prepared and measured after seven successive reductions in thickness from the original 0.032 in. to 0.0146 in. The curve φ_x in Fig. 5 gives the resultant graph of curvature *vs* distance to the original midsurface of the strip at $z=0$.

Equation (6) then requires the two further functions $d\varphi_x/dz_1$ and $\int_{z_1}^{z_0} \varphi_x dz$, both obtainable from this curve

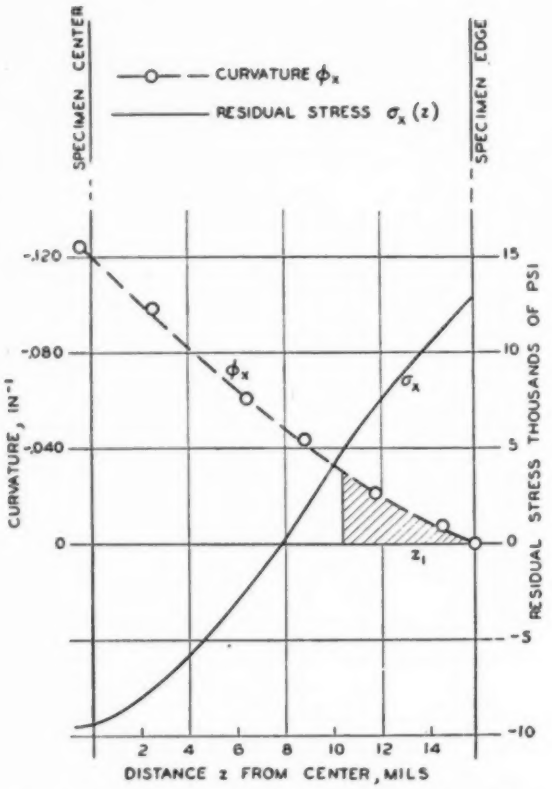


FIG. 5. Plot of measured curvature and calculated stress.

by graphical differentiation and integration at convenient intervals, not necessarily experimental points. For the abscissa point in the figure z_1, for example, the ordinate gives the value for $\varphi_x(z_1)$, the slope of the tangent the value of $d\varphi_x(z_1)/dz_1$, and the shaded area under the curve the value of $\int_{z_1}^{z_0} \varphi_x(z_1)dz$. These permit evaluation of $\sigma_x(z_1)$ by Eq. (6), and a suitable series of such values yields the complete stress distribution σ_x given in Fig. 5. Since the strip in this case was initially straight, the stress distribution could reasonably be assumed to be symmetrical about the midplane. Hence, the analysis was extended only over half the thickness. A check on the computed stress distribution is obtained by integrating the stress over half the thickness. By the condition of zero force this integral should vanish. The failure of the integral in Fig. 5 to vanish corresponds to an average error in stress of 770 psi or about 6 percent of the maximum. A symmetrical distribution necessarily satisfies the condition of zero moment.

The authors wish to thank Dr. R. D. Mindlin for valuable suggestions, particularly in regard to extending the results to the biaxial case, and Dr. W. C. Ellis for initiating and encouraging the investigation.

APPENDIX. SOLUTION OF THE INTEGRAL EQUATION

Since M_x is given, Eq. (4), in terms of known functions of z_1, Eq. (3) may be regarded as an integral equation

$$\int_{-z_0}^{z_1} \sigma_x(z)[z+(z_0-z_1)/2]dz = M_x(z_1), \tag{A1}$$

for the unknown stress σ_x in terms of the moment M_x. To obtain an explicit expression for $\sigma_x(z)$ we differentiate with respect to z_1, taking account of the fact that in the left-hand side z_1 occurs in both the upper limit and integrand. This gives

$$\sigma_x(z_1)\left[\frac{z_0+z_1}{2}\right]-\frac{1}{2}\int_{-z_0}^{z_1}\sigma_x(z)dz=\frac{dM_x(z_1)}{dz_1}. \tag{A2}$$

Rearranging Eq. (A2) and using Eq. (2) of the text gives

$$\sigma_x(z_1)=[1/(z_0+z_1)]\{F_x(z_1)+2[dM_x(z_1)/dz_1]\}. \tag{A3}$$

To express the force $F_x(z_1)$ in terms of the moment we differentiate Eq. (2) with respect to z_1, giving

$$dF_x(z_1)/dz_1=\sigma_x(z_1). \tag{A4}$$

Substituting into Eq. (A2), we have the following differential equation for F_x in terms of M_x:

$$(z_0+z_1)[dF_x(z_1)/dz_1]-F_x(z_1)=2[dM_x(z_1)/dz_1], \tag{A5}$$

or

$$(d/dz_1)[F_x(z_1)/(z_0+z_1)^2]=[2/(z_0+z_1)^2][dM_x(z_1)/dz_1].$$

Integrating from z_1 to z_0 and using the condition $F_x(z_0)=0$, Eq. (1),

$$\frac{F_x(z_1)}{z_0+z_1}=-\int_{z_1}^{z_0}\frac{2}{(z_0+z)^2}\frac{dM_x(z)dz}{dz}.$$

The right-hand side may be integrated by parts subject to $M_x(z_0)=0$, Eq. (1), giving

$$F_x(z_1)=\frac{2M_x(z_1)}{z_0+z_1}-4(z_0+z_1)\int_{z_1}^{z_0}\frac{M_x(z)}{(z_0+z)^3}dz. \tag{A6}$$

We now have an expression for the force F_x in terms of the moment M_x. To obtain a formula for σ_x involving only the moment M_x we substitute Eq. (A6) into Eq. (A3), giving

$$\sigma_x(z_1)=\frac{2}{z_0+z_1}\frac{dM_x(z_1)}{dz_1}+\frac{2M_x(z_1)}{(z_0+z_1)^2}-4\int_{z_1}^{z_0}\frac{M_x(z)}{(z_0+z)^3}dz. \tag{A7}$$

Substituting Eq. (4) for $M_x(z_1)$ and its derivative into Eq. (A7) gives the general stress-curvature relation (5) of the text which applies to an isotropic material.

For an anisotropic material the principal components of stress and strain in the xy plane are related by

$$e_x=s_{11}\sigma_x+s_{12}\sigma_y,$$
$$e_y=s_{12}\sigma_x+s_{22}\sigma_y.$$

If we let E_x and ν_x be the Young's modulus and Poisson's ratio determined by a uniaxial tension test with the tension in the x direction, then $s_{11}=1/E_x$ and $s_{12}/s_{11}=-\nu_x$. For a simple tension test with tension in the y direction the corresponding relations are $s_{22}=1/E_y$ and $s_{12}/s_{22}=-\nu_y$. Thus, $\nu_x/E_x=\nu_y/E_y$. In terms of the E's and ν's, Eq. (5) of the text takes the more general form

$$\sigma_x(z_1)=\frac{E_x}{6(1-\nu_x\nu_y)}\Big\{(z_0+z_1)^2\Big[\frac{d\varphi_x}{dz_1}(z_1)+\nu_x\frac{d\varphi_y}{dz_1}(z_1)\Big]$$
$$+4(z_0+z_1)[\varphi_x(z_1)+\nu_x\varphi_y(z_1)]$$
$$-2\int_{z_1}^{z_0}[\varphi_x(z)+\nu_x\varphi_y(z)]dz\Big\}, \tag{A8}$$

with a similar equation for σ_y obtained by interchanging the subscripts x and y.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

An Objective for Use in the Electron Microscopy of Ultra-Thin Sections

JAMES HILLIER
RCA Laboratories, Princeton, New Jersey and Sloan-Kettering Institute for Cancer Research, New York, New York
(Received August 31, 1950)

An objective has been designed which is particularly suitable for the study of tissue sections. It consists of a large opening polepiece system operated at a focal length of 12 mm. It includes a diaphragm which can be centered (not during operation) and translated on the axis. Angular apertures of as low as 2×10^{-3} radian with large areas of bright field can be achieved. A simple method of aligning the aperture is described.

I. INTRODUCTION

DURING the development of an ultra-thin sectioning technique in these laboratories, it was realized that it was not possible to remove the embedding materials by the use of conventional solvents without introducing artifacts which invalidated the use of the electron microscope.[1] The technique of using the beam to sublime off some of the embedding material in the vacuum of the microscope, overcame this difficulty but emphasized another, namely, the lack of contrast in specimens of the thicknesses involved. The use of the diaphragm which is standard for the RCA Model EMU Electron Microscope was not satisfactory for a number of reasons. Most important of these was the fact that the angular aperture of less than 5×10^{-3} radian which was usually desired was difficult to achieve and, in addition, involved a reduction of the bright field examination to an area less than only 25μ in diameter. An area at least 75μ and preferably 150μ in diameter appeared necessary for rapid orientation in the sections, that is, full bright fields at magnifications of 1000 and 500, respectively. Another difficulty was the result of the energy-spread in the imaging electrons transmitted by the thicker and denser portions of the sections. This introduced a linear diffusion of the image points owing largely to chromatic change in magnification. This required that accurate voltage alignment be achieved simultaneously with accurate centering of the diaphragm. Technically, this last requirement was very difficult to satisfy with the very small apertures that were necessary. The double objective lens described earlier[2] was tried. It provided the necessary contrast with a sufficiently large bright field but lacked the desired resolution because of its almost tenfold greater chromatic error.

It then became apparent that a polepiece designed especially for the examination of tissue sections would be desirable.

II. DESCRIPTION

A polepiece which fulfills most of the requirements indicated in the preceding section has been designed and has been found satisfactory after several months of routine use. A schematic cross section of the system is shown in Fig. 1. The construction is the same as that of the standard objective polepiece for the Model EMU, with the exception that the opening has been considerably enlarged (0.400 in.) and the gap increased (0.312 in.). The spacer was provided with compensating screws, but these have not been found necessary in any of the polepieces constructed. The diaphragm holder consists of tube and a cap. The largest possible diameters were used for this, though the diaphragm itself was only 0.125 in. in diameter, since a punch of that size was available. The tube is held by friction in a split ring which rests on the lower face of the polepiece gap. Two brass screws threaded through the spacer provide orthogonal centering motions. A single hairpin-shaped spring made of phosphor-bronze strip provides the return motions. For illustrative purposes one adjustment screw is shown opposite the spring. In the actual polepiece the screws are placed at 45° to each side of the point opposite the spring.

An adapter ring approximately 5 mm thick is used to raise the specimen holder for use with this lens. The lens then operates at a focal length of approximately 12 mm and at a magnification $\frac{1}{4}$ to $\frac{1}{5}$ that of the standard objective. The total magnification of the instrument using the highest power projector then covers the range 1500 to 5000 approximately. With the intermediate lens[3] the range is extended downwards to 200.

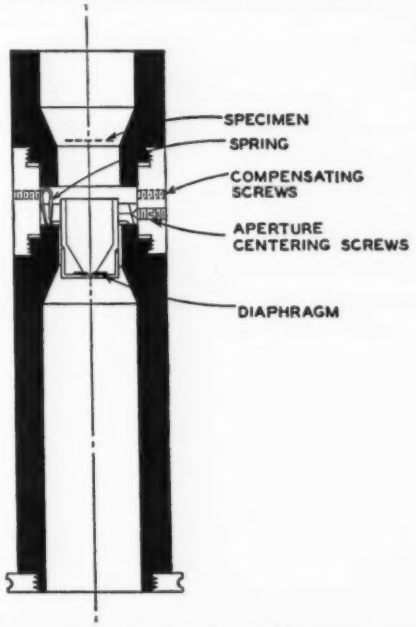


FIG. 1.

[1] J. Hillier and M. E. Gettner, J. Appl. Phys. **21**, 889 (1950).
[2] J. Hillier, J. Bact. **57**, 313–317 (1949).
[3] J. Hillier, J. Appl. Phys. **21**, 785–790 (1950).

To achieve an angular aperture of 10^{-2} radian requires a diagram approximately 240μ-diameter instead of the 50μ-diameter required for the standard lens. On the other hand, a diaphragm 50μ in diameter provides an angular aperture of 2×10^{-3} radian. In most cases the latter has been found to give excessive contrast for the photographic materials commonly used.

When, as in this lens, the diaphragm is placed in the image space, its axial position exerts little influence on the angular aperture if the separation of the lens and the diaphragm is small compared with the image distance. On the other hand, the axial position as well as the diameter of the diaphragm can be the primary factors determining the diameter of the bright field region. The bright field region is defined by those rays which pass through the specimen *without* being scattered. This means that the bright field region is determined entirely by the optics of the illuminating system and the objective. Under these conditions the exit pupil of the objective is the image produced by it of the condenser diaphragm. Regardless of the adjustment of the illuminating system, none of the bright field rays can pass outside the exit pupil. Thus, an objective diaphragm which has a diameter *larger* than this exit pupil and is placed at the same axial position can never restrict the bright field region. With the new lens and the standard condenser diaphragm, the exit pupil has a diameter of approximately 43μ and is situated approximately 13 mm below the central plane of the lens. Thus, a diaphragm of 50μ-diameter or larger placed at the plane of the exit pupil and accurately centered cannot limit the bright field region. Furthermore, such a diaphragm will never receive any direct bombardment.

The diaphragms used were made from 125μ-thick copper. A fine sewing needle which, was sharpened on the finest available stone so that its tip consisted of a triangular pyramid with a point less than 10μ in diameter, was used as a drill. The drilling was done by hand with the copper placed on a glass plate. In this arrangement a smooth hole less than 25μ in diameter was produced. This was then etched carefully in nitric acid until the opening had the desired diameter. In addition to providing a control over the size of the aperture, the etching effectively removes small particles of the stone which are transfered on the needle and are embedded in the walls of the hole. A few companies have made diaphragms of platinum available to electron microscopists. Unfortunately, in the tests made in this laboratory, the proportion which was "electron optically" clean was so small that they cannot be recommended.

The complete adjustment of the new objective and its diaphragm involves three operations: (1) accurate voltage alignment[4] of the entire microscope, at the maximum magnification to be used, and without the objective diaphragm; (2) lateral adjustment (centering) of the diaphragm; (3) axial adjustment of the diaphragm to place it at, or slightly below, the plane of the image of the condenser diaphragm. Adjustment after the replacement of the diaphragm required only the first two operations.

[4] J. Hillier and E. G. Ramberg, J. Appl. Phys. 18, 48–71 (1947).

Centering of Diaphragm

The following procedure appeared to give the most direct centering of the objective diaphragm in this lens. The instrument was set up with the new objective inserted in a definite angular position with the adapter on the specimen holder and with the projection lens arrangement to be used in the future work. The objective diaphragm was omitted in this first adjustment. With this arrangement accurate voltage alignment was achieved and the alignment adjustments tightly clamped. The diaphragm was then inserted (centered by eye) being careful to maintain the same angular position of the objective. The instrument was then operated in the following way: The projection lens polepiece and the intermediate lens diaphragm were removed and the projection lens current turned off. The condenser current was turned to its maximum value. The objective current was set somewhat above the focal value. With this arrangement of the instrument, the shadow of the diaphragm, as observed on the final screen, is erect and not rotated by the field of any lenses. This means that a motion of the diaphragm in any direction as observed from above will produce a motion of its shadow in exactly the same direction. The amount of motion necessary to center the diaphragm was determined by the amount, in terms of the known diameter, that the shadow was off center. Thus, this arrangement gave simple criteria which made it possible to center the diaphragm accurately and quickly. After this system was devised, more than two trials were rarely necessary. The diaphragm was considered accurately centered when the bright field had infinite extent for the focal value of the objective current and when its shadow contracted symmetrically as the objective current was varied about the focal value and during the early stages of change when the accelerating potential was turned off. This test is significant only if the axial position of the diaphragm has been adjusted.

Axial Adjustment of the Diaphragm

On the basis of first-order theory, the calculated position of the diaphragm is approximately 9.2 mm below the lower face of the polepiece gap. However, a number of experimental factors can change this value somewhat, so it is desirable to have an experimental method of making the adjustment accurately. With the exception of leaving the value of the objective current fixed, the arrangement of the instrument used for this adjustment was identical with that given above. The diaphragm was inserted so that it was about 9.5 mm below the lower surface of the supporting ring and

FIG. 2. Electron micrographs of the same field of view of a 0.2μ-section of mouse liver. Embedding partly removed by bombardment. (Top) New lens, no diaphragm. Printed for maximum contrast. (Middle) New lens with 182μ-aperture. Angular aperture 7×10^{-3} radian. (Bottom) New lens with 61μ-aperture. Angular aperture 2.3×10^{-3} radian. Printed for medium contrast. Printing identical with middle picture.

centered according to the above procedure. Then the behavior of the shadow of the diaphragm was observed as the value of the objective current was raised and lowered. The value for which the shadow appeared to expand to infinity was observed. It was necessary to interpolate here, since there was an appreciable range within which the edge of the shadow was outside the image field. This interpolated value was compared with the previously determined focal value of the objective current. If it was higher, the diaphragm needed to be lowered and vice versa. The indicated adjustment was made until the two values coincided. It was important that the condenser current be maintained at its maximum value during these adjustments.

Theoretically, this method of adjustment places the diaphragm about 0.5 mm below the image of the condenser diaphragm. This was considered a good practical adjustment, since for the smallest diaphragms (50μ-diameter) the theoretical tolerances on the adjustment are approximately 1 mm on the lower side and only 0.1 mm on the higher side.

IV. DISCUSSION

Three lenses of this type have been constructed and tested. Their performance has been remarkably similar and in good agreement with that expected from the discussions given above. Since, in these lenses the only bombardment received by the diaphragm is from the electrons scattered by the specimen, the build-up of the contamination and the resulting appearance of image asymmetries takes place much more slowly than in the conventional lens. One diaphragm was used routinely for three months before serious asymmetries appeared. In order to achieve this, however, it was obvious that the adjustments must be made accurately and maintained. If, through subsequent misalignment of the illuminating system, or through improper centering of the diaphragm, it was allowed to receive direct bombardment, large scale asymmetries appeared almost immediately. Two aspects of the optics of the lens are responsible for this. In the region of the objective lens, the direct electron beam has its maximum intensity near the level of the diaphragm. This means that the contamination rate is very high if this beam or part of it strikes the diaphragm. At the same time the beam at the level of the diaphragm has its minimum diameter. Since that minimum diameter is much smaller than the diameter of the smallest diaphragm used, any bombardment of the diaphragm and the resulting contamination must be very asymmetrical. These observations have emphasized the need for accurate adjustment of the lens and the resulting benefits.

Figure 2 illustrates, as well as is possible through reproduction, the effect of different apertures on the contrast obtainable.

ACKNOWLEDGMENT

The author wishes to acknowledge that, in the course of work on another project, S. G. Ellis made a number of observations which helped to clarify the behavior of this new lens.

The Leaky-Condenser Oxide Cathode Interface*

A. EISENSTEIN
Department of Physics, University of Missouri, Columbia, Missouri
(Received August 11, 1950)

The time dependence of current passed by commercial pentodes has been studied under pulsed conditions of operation. A decay of current is observed in the microsecond time range for those tubes operated for long periods under cut-off conditions. This decay has been interpreted in terms of the resistance and capacitance of the interface layer known to exist from x-ray diffraction studies. Techniques are developed and applied to the measurement of this interface resistance and capacitance as a function of temperature. Following the initial current decay, a partial recovery is observed when long pulses are used. This effect results from Joule heating of the interface layer.

ALTHOUGH the oxide cathode has been the subject of study during the past fifty years, the presence of the interface layer was not generally recognized before 1940. At one time it was hoped that many unexplained cathode phenomena could be interpreted in terms of this new parameter. This was a vain hope, for we now know only a few cases in which the electronic properties of the cathode can be correlated with the presence of the interface. All of these appear only under special conditions of cathode operation and are not of particular significance in the conventional dc operation of cathodes.

Figure 1(A) shows an enlarged cross-section view of the cathode with the interface layer sandwiched between the base metal and the oxide coating. This interface compound is formed in a solid-state chemical reaction between the oxide and the base metal or with reducing constituents of the base metal, e.g., Si, Ti, Al, etc.[1] When the base metal is a Si-Ni alloy, the interface compound is barium orthosilicate, Ba_2SiO_4. Since electrons emitted from the vacuum surface of the cathode must traverse both the coating and interface, the electronic properties of both may influence the operating characteristics of the cathode. Rectification has been shown[2] to occur at the base metal-semiconductor junction and to be much enhanced when the Ba_2SiO_4 interface is present. The high specific resistivity of this interface, as compared to the coating, may give rise to a relatively large voltage drop when current is drawn from the cathode. When this voltage drop becomes an appreciable fraction of the applied anode-cathode voltage, an apparent deviation from the theoretical diode voltage-current characteristic is found.[3] At still higher emission currents, the interface voltage drop may actually limit the maximum available current through the phenomena of "sparking." These properties arise due to the resistive effect of the interface layer.

Other properties of the cathode showing a time dependence may be explained in terms of the resistive-capacitive effect of the interface. The reported[4] variation of transconductance with frequency and with tube life, as well as the decay of current passed by vacuum tubes with decay constants in the microsecond time range can be so explained. Some commercial tubes develop this decay after several thousand hours life under cut-off conditions of operation, i.e., heater power applied but no emission current drawn. This decay is responsible for tube failure in many electronic computer applications. The author was introduced to the problem by the group at the M.I.T. Servomechanisms Laboratory, and he wishes to thank particularly E. S. Rich of that Laboratory who supplied most of the tubes used in this study.

THEORETICAL CONSIDERATIONS

The cross-section view of a diode, Fig. 1(A), can be represented by an equivalent circuit diagram. R_1 represents the interface resistance, R_2 the coating resistance, R_3 the effective diode resistance, and R_4 the series external resistance. The parallel interface capacitance C is shown. Although a parallel coating capacitance and an interelectrode capacitance also exist, they are usually negligible with respect to C. When an external voltage V_p is suddenly applied to the diode, a total initial current i_0 will flow but will rapidly decrease to the steady state value i_s, Fig. 1(B). This total current represents the sum of two currents, i_1 through the resistance R_1, and i_2 through the capacitor C. The time dependence of these currents is,

$$i_1=\frac{V_p}{R_1+R_2+R_3+R_4}\times\left[1-\exp\left(-\frac{R_1+R_2+R_3+R_4}{R_1C(R_2+R_3+R_4)}t\right)\right], \quad (1)$$

$$i_2=\frac{V_p}{R_2+R_3+R_4}\exp\left(-\frac{R_1+R_2+R_3+R_4}{R_1C(R_2+R_3+R_4)}t\right), \quad (2)$$

and

$$i_0=\frac{V_p}{R_2+R_3+R_4}, \quad (3)$$

$$i_s=\frac{V_p}{R_1+R_2+R_3+R_4}. \quad (4)$$

* This work supported in part by the ONR.
[1] A. Eisenstein, *Advances in Electronics* (Academic Press, New York, 1948), Vol. I, p. 24.
[2] W. E. Mutter, Phys. Rev. **72**, 531A (1947).
[3] A. Eisenstein, J. Appl. Phys. **20**, 776 (1949).
[4] W. Raudorf, Wireless Engineer October (1949).

Thus the extent of the current decay i_0/i_s depends on the relative value of R_1 with respect to the other series resistances; only the rate of current decay depends on the value of C.

This model and these relationships were originated by Wright[5] to show that this mechanism for decay was *not* effective in producing a decay which he observed to have a decay time constant in the millisecond range. By making assumptions of the interface thickness and dielectric constant to evaluate C, and by using measured values of R_1, Wright concluded that the decay time constant D for this model would be less than one microsecond,

$$D=\frac{R_1C(R_2+R_3+R_4)}{R_1+R_2+R_3+R_4}. \tag{5}$$

Although inappropriate for Wright's observed millisecond decay, this mechanism will be applied to the problem of current decay in certain tubes whose time constant falls in the microsecond time range.

In diodes, the effective impedance R_3 is generally large compared to the interface resistance R_1, the ratio i_0/i_s approaches unity, and the decay effect is small. In multielement tubes, however, the effect is magnified because of the voltage amplification of the tube. As shown in Fig. 2, the parallel R_1C of the interface appears in series with the cathode lead, thus any voltage developed across this element appears in both the grid and plate circuits. The electrical behavior is thus that of a degenerative amplifier and may be treated by the mathematical analysis that has been applied[6] to this well-known case. A fixed negative bias V_g is applied to the grid in series with a crystal rectifier which prevents the grid from being driven positive. When a positive going voltage pulse e_g is applied to the grid, an output current pulse i_p appears in the plate circuit. This pulse has the form,

$$i_p=e_gg_mf(t), \tag{6}$$

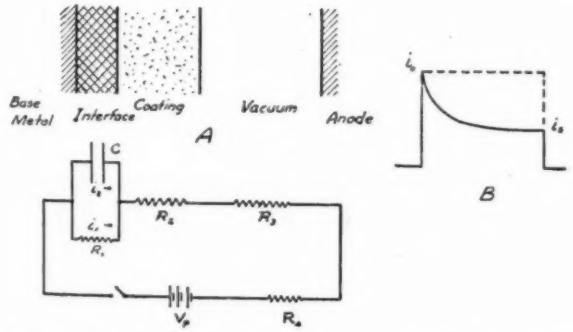


FIG. 1. (A) Enlarged cross-section view of simple diode and its equivalent circuit diagram; (B) Pulse shape showing decay from initial current i_0 to steady-state current i_s.

[5] D. A. Wright, Proc. Phys. Soc. (London), Sec. B **62**, 188 (1949).

[6] G. E. Valley and H. Wallman, *Vacuum Tube Amplifiers* (M.I.T. Radiation Laboratory Series, McGraw-Hill Book Company, Inc., New York, 1948), p. 87.

where g_m, the mutual conductance, is defined as the slope of the E_g-i_p tube characteristic $(\Delta i_p/\Delta E_g)$ and $f(t)$ is a time dependent function, which for the degenerative amplifier is,[6]

$$f(t)=\frac{g_mR_1}{1+g_mR_1}\left[\frac{1}{g_mR_1}+\exp\left(-\frac{1+g_mR_1}{R_1C}t\right)\right]. \tag{7}$$

Supplying $t=0$ and $t=\infty$ values leads to expressions for the initial current i_0 and the steady-state current i_s,

$$i_0=e_gg_m \tag{8}$$

$$i_s=e_gg_m\frac{1}{1+g_mR_1}. \tag{9}$$

Combining these gives,

$$R_1=\frac{i_0/i_s-1}{g_m}. \tag{10}$$

The decay time constant for this case is,

$$D=R_1C/(1+g_mR_1). \tag{11}$$

A second problem which is of interest and is related to the decay effect is the recovery of current after the end of the pulse. For example (Fig. 3), following the end of a pulse of duration t_1, what current would be passed by the tube at a time t_2? Mathematically this is a simpler problem, since after the pulse the interface capacitance can discharge only through its own internal resistance R_1, thus behaving as a leaky condenser. The interface voltage e_i decreases exponentially with time from its value at the end of the pulse. When the pulse is of sufficient duration that the steady-state plate current i_s is reached,

$$e_i=i_sR_1\exp-(t_2-t_1)/R_1C. \tag{12}$$

Upon reapplication of the grid pulse e_g, the initial grid potential is (e_g-e_i). The current $(i_p)_{t_2}$, which would be passed by the tube at time t_2 upon the re-application of the grid pulse e_g, is thus $g_m(e_g-e_i)$ and making use of Eqs. (12) and (9),

$$(i_p)_{t_2}=e_gg_m\left\{1-g_mR_1\frac{\exp-[(t_2-t_1)/R_1C]}{1+g_mR_1}\right\}. \tag{13}$$

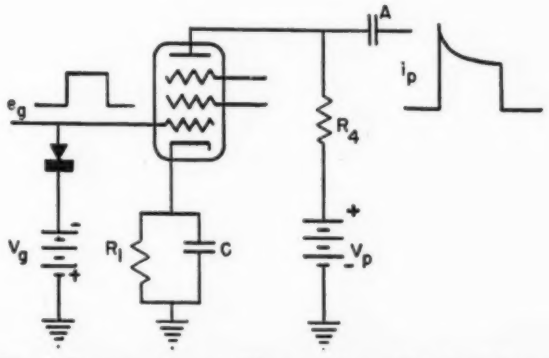


FIG. 2. Degenerative amplifier circuit resulting from interface resistance and capacitance in cathode circuit. Input grid pulse e_g and output current pulse i_p appearing at A.

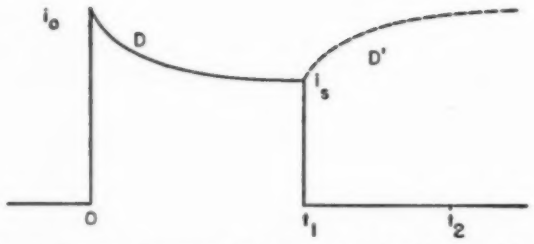


FIG. 3. Decay of current during pulse of duration t_1 with time constant D. Dash line indicates recovery of current to the initial value i_0 with time constant D'. This curve represents the current which would be available if a second pulse were applied, say at time t_2.

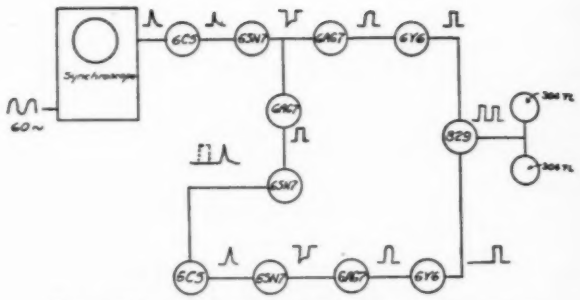


FIG. 4. Diagram showing tube line-up and operating principle of double pulse generator.

This current recovers exponentially from i_s to the initial value i_0, with a recovery time constant D',

$$D'=R_1C. \tag{14}$$

Combining this expression with (11),

$$R_1=[(D'/D)-1]/g_m, \tag{15}$$

and comparing with (10),

$$D'/D=i_0/i_s. \tag{16}$$

The leaky condenser interface model affords several checks on its validity. The ratio D'/D may be measured and compared with i_0/i_s; R_1 and C may each be measured as a function of temperature to determine if the variation observed is that to be expected of the model.

Raudorf[4] has proposed a similar leaky condenser interface model on the basis of the measured variation of g_m with frequency in the range 10^5 to 10^7 cps for pentode type tubes subjected to long life test periods. His model differs only in mechanical detail from that proposed here, but significantly in that the predicted operating characteristics of Wright's model should bear a marked dependence on the base metal impurity present, whereas this is not necessary in Raudorf's model.

EXPERIMENTAL RESULTS

I. Microsecond Pulse Decay and Recovery

In order to observe the time dependence of current in pentodes, the basic circuit of Fig. 2 was used. The

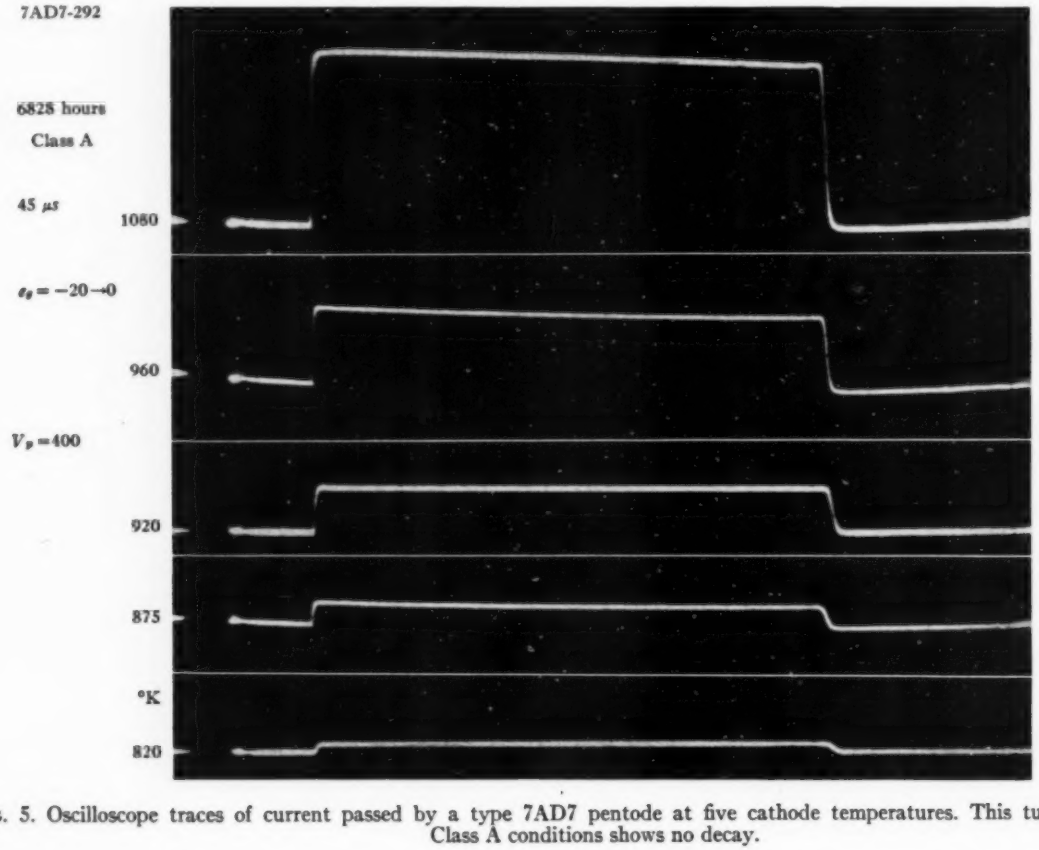


FIG. 5. Oscilloscope traces of current passed by a type 7AD7 pentode at five cathode temperatures. This tube operated under Class A conditions shows no decay.

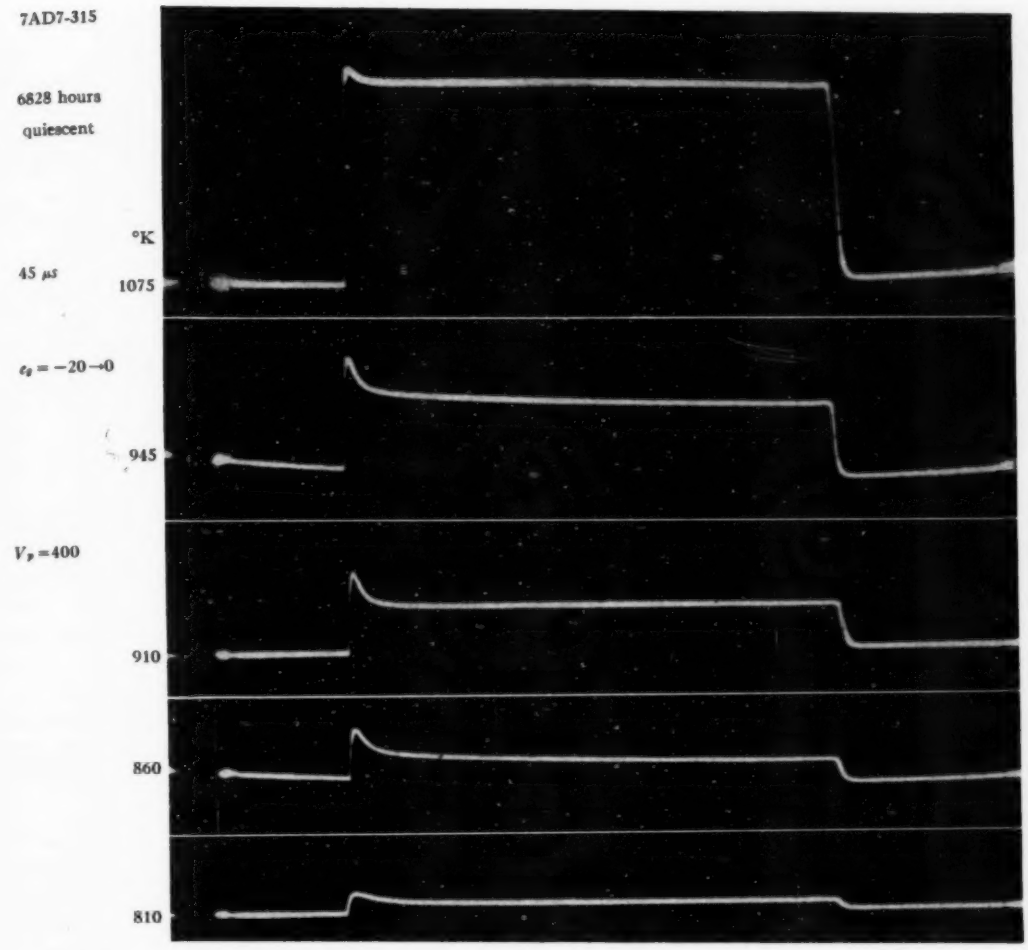


FIG. 6. Oscilloscope traces of current passed by a type 7AD7 pentode at five cathode temperatures. This tube operated under quiescent (cut-off) conditions shows the decay effect.

plate resistance R_4 was either 500 or 1000 ohms and the variation of current was determined by observing the potential at point A with respect to ground. Both the Sylvania Model P-4 synchroscope and the Tektronix Model 511-AD oscilloscope were used in this study. The grid pulse was supplied from a special double-pulse generator† indicated in Fig. 4.

Pentode tubes used in this study were supplied by the M.I.T. Servomechanisms Laboratory, and were commercial type 7AD7 tubes. The base metal used in these tubes contained about 0.1 percent silicon. Three groups of tubes have been investigated, (I) new tubes, as received from the manufacturer, (II) tubes operated more than 5000 hours under quiescent conditions, i.e., biased beyond cutoff so that no emission current flowed, and (III) tubes operated more than 5000 hours under class A conditions, i.e., normal rated emission current being drawn.

As found by others,[7] only tubes in group (II) exhibit the decay characteristic. Figure 5 shows a set of five oscilloscope traces of the plate current response of a 7AD7 which had been subjected to 6828 hours of operation under class A (group III) conditions. This tube, operated with 400 volts applied to the plate and screen, was driven from near cutoff, −20 volts, to zero bias

† In the initial work to be described, only single pulses at a repetition rate of 60 pps were used, however, in later work, two pulses, each variable both in duration and in spacing, were used. The synchroscope is locked in with a 60-cycle input signal and the trigger pulse developed is introduced into a 6SN7 multivibrator through the 6C5 isolating amplifier. Two 6AG7 tubes differentiate and partially square the multivibrator output pulse. One of these is introduced into a 6Y6 cathode follower for further shaping and finally into one-half of an 829. The other pulse serves to trigger a 6SN7 one-shot multivibrator whose output is delayed by a variable time length. This delayed pulse is introduced as a trigger for the second pulse channel whose operation is identical to the first. Negative going double pulses are taken from the two plates of the 829 which have a common plate load resistor. These pulses drive a pair of 304TL triodes, connected in parallel and operating at zero bias, to beyond cutoff. This produces positive going output pulses having very flat tops and negligible "droop" with the 4 mf output condenser and a 800-ohm plate load resistor used. The output stage operates at a maximum average current of about 1 ampere at 1500 volts and the output pulse is variable in amplitude between 0 and 750 volts. Since positive going driving pulses of only about 25 volts were required for the pentodes used in this work, a step down resistance divider was used.

[7] E. S. Rich (private communication). R. M. Matheson and L. S. Nergaard, Phys. Rev. **78**, 355(A) (1950).

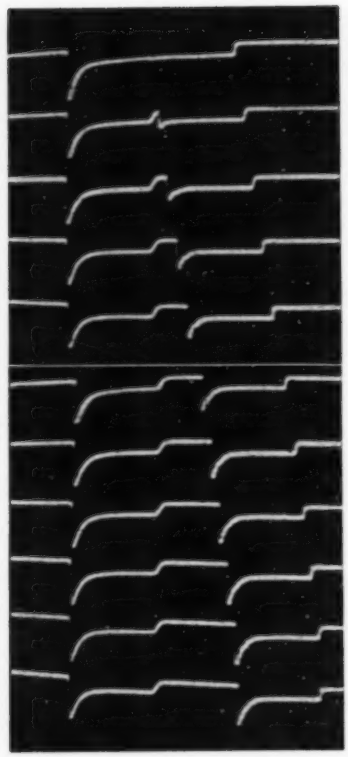


FIG. 7. Double current pulses passed by 7AD7 pentode as the delay between pulses is varied. These traces are inverted, i.e., a downward deflection corresponds to an increase of current.

with a 45-microsecond grid pulse. Each curve represents the response at a different cathode temperature. These temperatures were determined from heater wattage measurements through the use of a wattage *vs* cathode base metal temperature chart. This chart was determined for one tube having a Mo-Ni thermocouple attached to the center of the cathode under the coating. This calibration tube was made of 7AD7 tube parts obtained by dismantling one of the tubes. Figure 6 shows a similar set of five oscilloscope traces of the plate response of a 7AD7 also subjected to 6828 hours life test, but under cut-off conditions (group II). The decay effect is apparent and the ratio i_0/i_s is seen to be temperature dependent, increasing as the temperature is reduced.

In order to study the recovery time constant D' as well as the decay constant D, the double pulse technique was used. Figure 7 shows a series of oscilloscope traces of the plate current of the pentode during a pair of 3-microsecond pulses, as the interval between the pulses is varied from 0 to 3 microseconds. These traces are inverted, i.e., a downward deflection corresponds to an increase in current. The current available at the beginning of the second pulse is seen to increase with the duration of the delay between pulses starting from i_s and increasing to i_0. To facilitate the determination of the time constants D and D', a procedure was adopted for superimposing a set of traces with differing delays, such as in Fig. 7, to form a single photograph. This was accomplished by using the double exposure technique to record, on a single film in the oscilloscope camera, traces of the initial and the delayed pulse, the latter being recorded with up to 20 different delay times. A typical set of these superimposed traces are shown in Fig. 8 for a cathode at four temperatures. The recovery of current passed by the tube is seen to increase from i_s to i_0 but always requiring a considerably longer time than did the decay. The envelope formed by the "spikes" at the beginning of the delayed pulse can be used to evaluate the recovery time constant D' in the same manner that the shape of the initial pulse can be used to determine the decay constant D. In this work, these constants were found by first making an optical projection of the photographic negative onto graph paper and tracing the enlarged image of the decay curve and the envelope of the recovery curve. Measurements of i_0 and i_s were also made at this time. Values taken from the two graphed curves were used to construct semilog plots of $(i_p - i_s)$ *vs* time. In over one hundred plots constructed in this manner, it was found that a straight line would best represent the points thus indicating an exponential time dependence as predicted by Eqs. (6) and (13). The slopes of these semilog plots were used to evaluate D and D' directly in units of microseconds, these representing the times required for the currents to change by $1/e$.

Table I shows a complete analysis made for a 7AD7 pentode following the procedure outlined. From the ratio D'/D, R_1 was determined, using Eq. (15) once the mutual conductance, g_m, and its temperature dependence was known. These values were determined from the slope of the $i_p - E_g$ characteristic and obtained under pulsed conditions. A typical set of these curves is seen in Fig. 9. Due to the difficulty experienced in operating the tube at a fixed bias V_g and setting the grid pulse amplitude e_g to the desired value, the procedure adopted was that of setting a fixed value of e_g (say +25 volts) and varying the bias V_g, since the effective grid voltage E_g is the sum of V_g and e_g. The mutual conductance is seen to vary from 1460 to 4000 micromhos in the temperature range 830°K to 1085°K. It is also seen that due to temperature limited emission, the curves drop below the dashed lines and are thus no longer represented by a single value of g_m as E_g approaches zero grid volts. For this reason, data taken for use in measuring R_1, e.g., that in Table I, is confined to E_g values more negative than −6 volts such that a single value of g_m represents the operating characteristics of the pentode. It should be noted that values of g_m and their temperature dependence differ considerably from tube to tube and must be measured for each. Having determined D' and R_1, C was found from Eq. (14).

Figure 10 shows the temperature dependence of D and D' as found in two tubes. It is noted that D is less than D', and shows a smaller temperature dependence

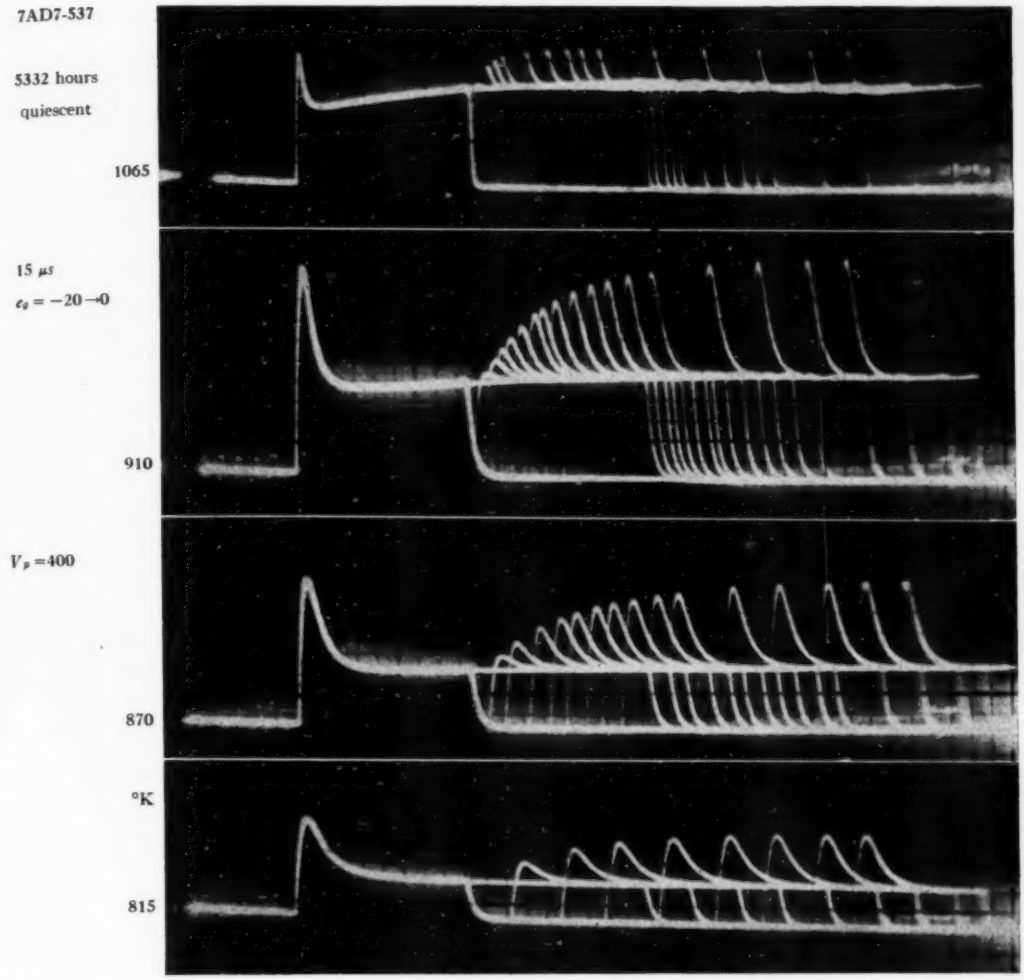


FIG. 8. Decay and recovery of current in pentode at four cathode temperatures. Double pulse technique used with variable delay between pulses. Initial pulse 15μs duration.

in accord with the predictions of Eqs. (11) and (14). The temperature dependence of D' is due primarily to the changing interface resistance, R_1. This effect is largely cancelled in (11) by the inverse variation of g_m with temperature to cause D to remain nearly constant. It is observed that the ratio D'/D is nearly double the ratio i_0/i_s in disagreement with the predictions of Eq. (16). This discrepancy is due in part to an incorrect measurement of i_0, the zero time current. The pulser used had a rise time of 0.2μs, hence, oscilloscope measurements gave a value less than the true i_0. For this reason Eq. (15) rather than Eq. (10) was used for evaluating R_1. The temperature dependence of R_1 is shown in Fig. 11 in which $\log R_1$ is plotted as a function of $1/T$. The linear variation of the function in this type of plot gives some assurance that the resistance measured is that of a semiconductor. A conductivity activation energy of 0.8 ev is determined from the slope. Other interface resistances measured have yielded similar values of Q, all lying in the vicinity of 1 ev.

The interface capacitance C is seen, Table I, to be essentially independent of temperature. On the basis of the interface model used to describe the decay effect, an estimate of the interface thickness can be made. The capacitance of a plane parallel plate condenser is given by,

$$C_{\mu\mu f} = 0.0885kA/t, \tag{17}$$

where k is the dielectric constant, assumed here to be 10, A is the area, 1.5 cm^2 for the cathodes used, and t is the condenser spacing. Using the measured value 1.9×10^4 $\mu\mu f$ for C, t is 7×10^{-5} cm. This thickness is consistent in magnitude with observed[3] interface thicknesses found in Si-Ni alloy base cathodes. On the basis of the calculated thickness and the observed resistance at 970°K, $R_1 = 310\Omega$, the effective specific electrical conductivity may be calculated to be 1.5×10^{-7} Ω^{-1} cm^{-1}. Here again this value agrees in order of magnitude with the observed[3] specific conductivity of Ba_2SiO_4. It is noted, however, that the conductivity of this silicate, as with other semiconductors, is subject to wide variation depending upon the degree of activation.

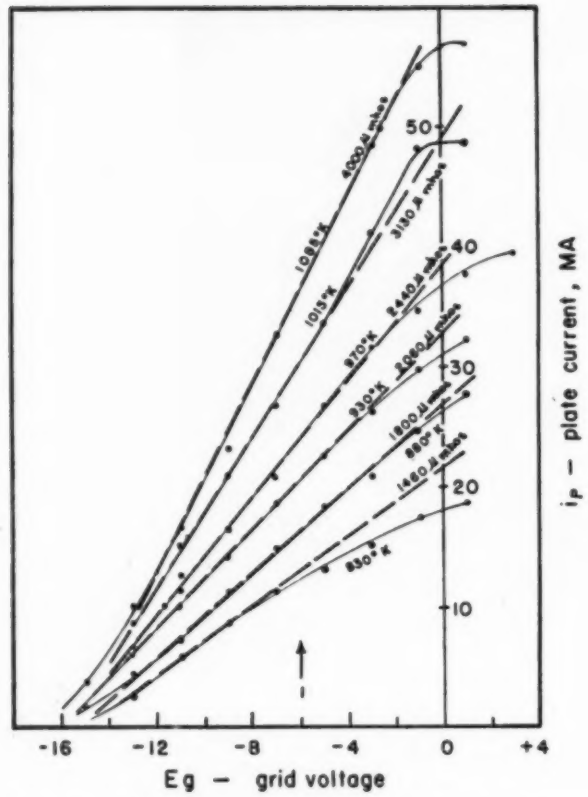


FIG. 9. Typical $i_p - E_g$ characteristic for 7AD7 for several cathode temperatures. Values of g_m shown refer to straight, dash lines matched to experimental, curves.

In order to definitely confirm the presence of the Ba_2SiO_4 interface in cathodes showing the decay, x-ray diffraction patterns were taken of cathodes removed from pentodes following the recording of their electrical behavior. The oxide coating was removed by dipping in acetone and exposing the interface region for x-ray diffraction analysis. Figures 12(A) and (B) are patterns of cathodes from group I. In these patterns of new cathodes only the diffraction lines of the nickel base metal are seen. Patterns C and D of group II cathodes snow the presence of the Ba_2SiO_4 interface as evidenced by the group of weaker diffraction lines appearing near the center. Absence of the interface in the group I

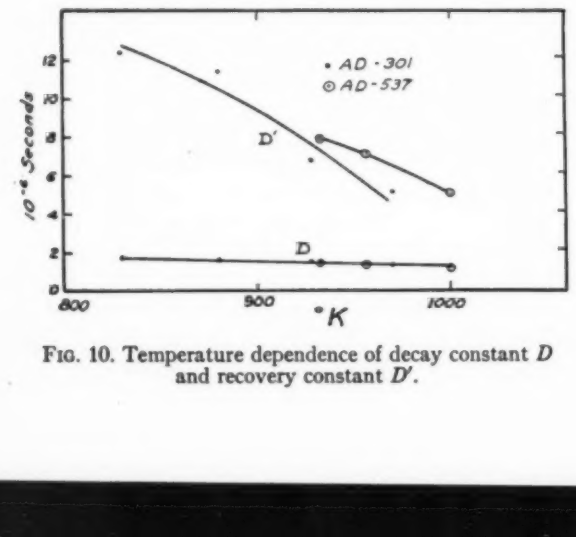


FIG. 10. Temperature dependence of decay constant D and recovery constant D'.

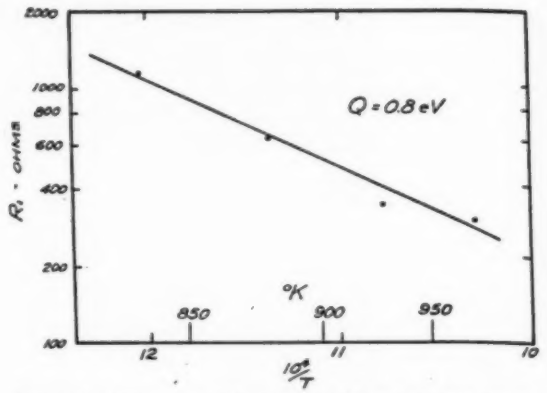


FIG. 11. Temperature dependence of interface resistance. Activation energy determined from slope.

cathodes showing no decay, and its presence in the group II cathodes, which exhibit the decay, is taken as evidence that the interface model assumed here is correct. From the relative intensities of the nickel base metal lines and the interface lines, it is estimated that

TABLE I.

	Tube AD-301 Type 7AD7 V_p = 400 volts E_g = −20 → −6					Life conditions 5577 hours quiescent		
V_f volts	T °K	$D\times10^4$ sec	$D'\times10^4$ sec	D'/D	i_s/i_s	g_m μmho	R_i ohms	C mfd
3.5	830	1.73	12.4	7.2	3.3	5500	1130	0.011
4.0	880	1.59	11.4	6.6	3.1	8300	603	0.019
4.5	930	1.46	6.5	4.4	2.2	10,000	340	0.019
5.0	970	1.29	5.1	4.0	1.9	10,200	294	0.017
6.3	1085	1.16	3.6	3.1	1.2	10,500	200	0.008

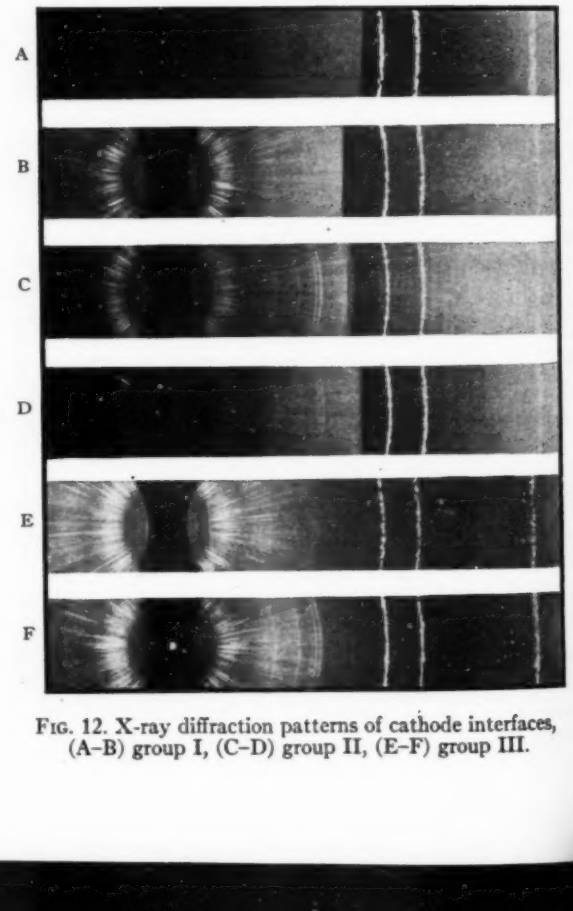


FIG. 12. X-ray diffraction patterns of cathode interfaces, (A–B) group I, (C–D) group II, (E–F) group III.

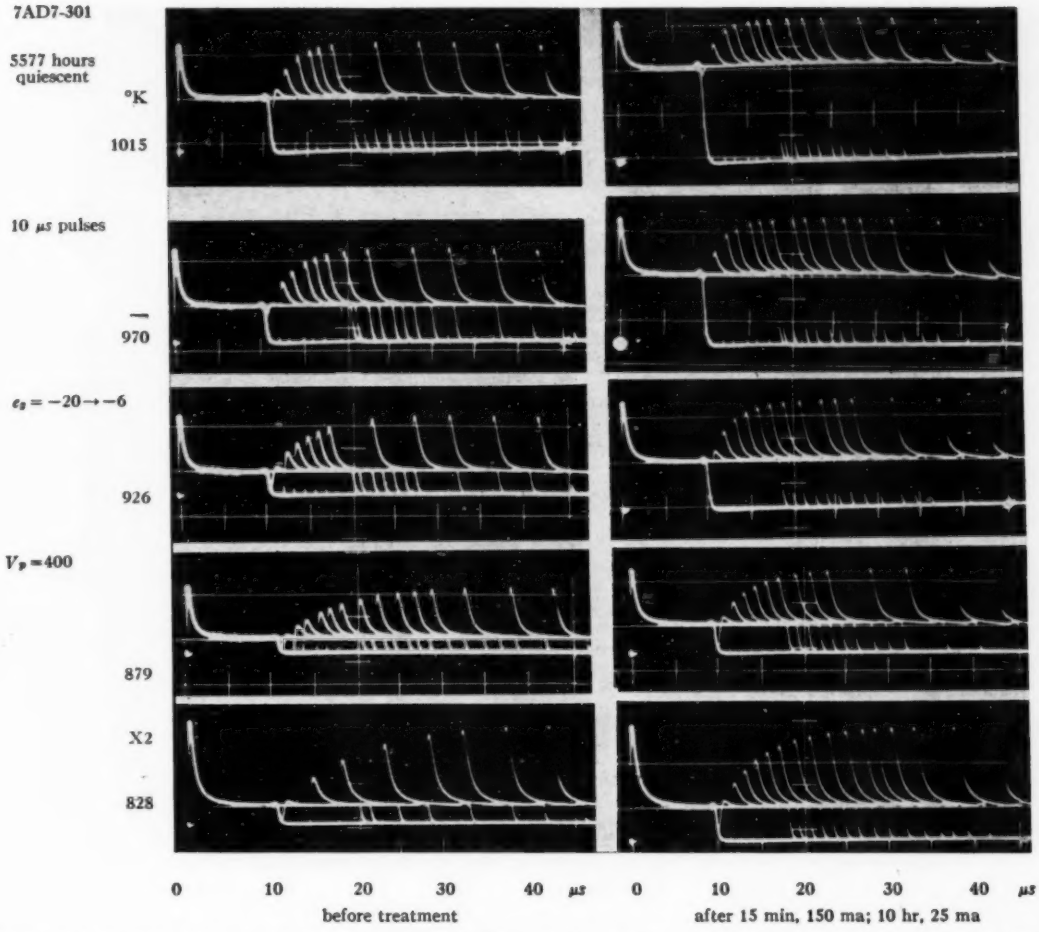


FIG. 13. Decay and recovery of current pulses passed by 7AD7 pentode at five operating temperatures before and after activating treatment. Note reduced ratio i_0/i_s after activation.

a value of 10^{-4} cm expresses the order of magnitude of the interface thickness. Also consistent with this interface model is the observation[7] that tubes containing no silicon in the cathode base metal do not develop the decay under group II operating conditions.

With this direct correlation of the presence of the interface compound and the decay phenomena a question arises concerning the absence of the decay in group III cathodes, which have been subjected to the same heat treatment as those of group II exhibiting decay. Patterns E and F of Fig. 12 are of the interfaces of cathodes from group III; the presence of a Ba_2SiO_4 interface of similar thickness is evident. Although this might seem to be inconsistent with the aforementioned correlation of the interface and the decay, it must be considered that the interfaces of the group II and group III cathodes were subjected to differing electrical conditions during their formation. Previous measurements[3] have demonstrated that the state of activation of Ba_2SiO_4 can be increased appreciably, and thus also its specific conductivity, by the flow of an electron current. Ionic migration of barium in the coating to the interface region would be considerably enhanced in the group III cathodes operated under class A conditions during the interface formation process. This excess barium may further activate the interface. On the basis of these hypotheses it is not surprising that the interfaces of group III cathodes show a much lower resistance than those of group II.

If the interfaces of the group II cathodes are merely in a lower activation state than those in group III, it should be possible to effect some decrease in the interface resistance by activation due to electron current flow. Some degree of success has been found in this procedure but in no cases has the interface resistance been eliminated. Figure 13 shows a set of current pulses passed by a group II pentode, before treatment and after being subjected to a dc activation, by drawing $0.1A/cm^2$ for 15 minutes followed by a stabilizing current density of $0.017A/cm^2$ for 10 hours. Before activation the conductivity, activation energy was 1.0 ev, and after activation it was 0.7 ev. The effect of this activation treatment is most pronounced at high temperatures where the interface resistance was reduced by a factor of two. This can be seen in the figure by comparing the ratio i_0/i_s which is related to the interface resistivity.

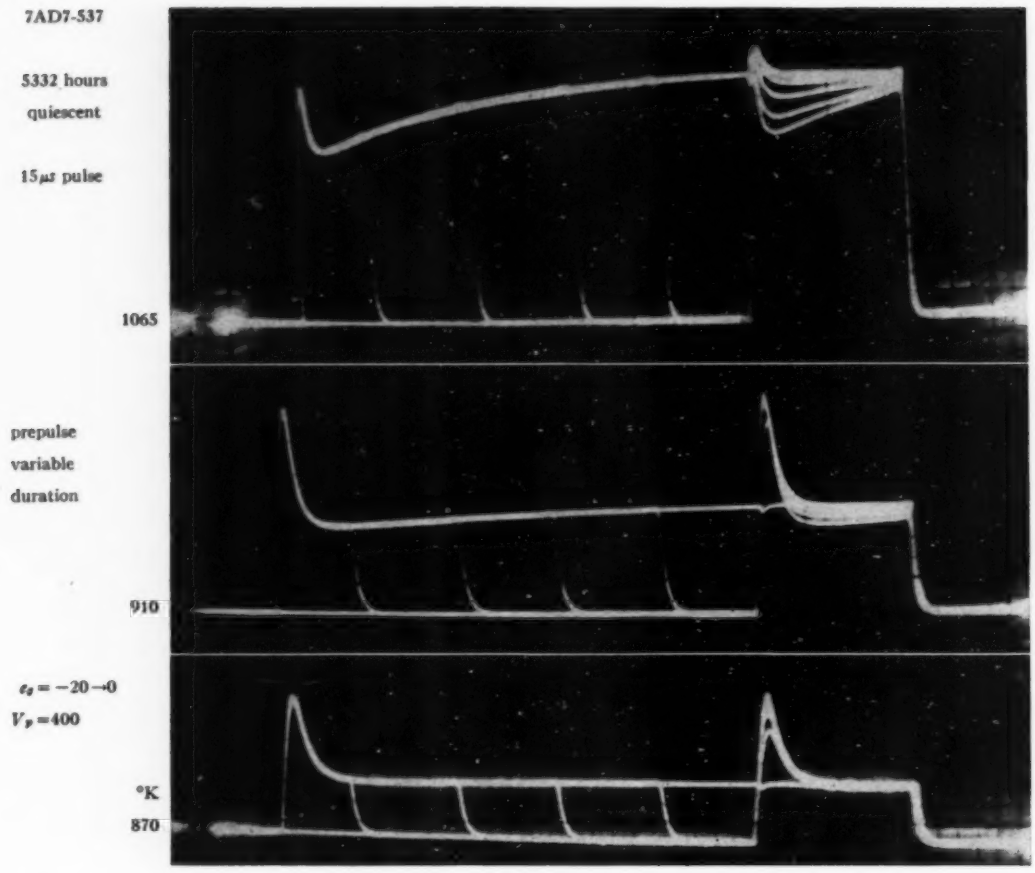


FIG. 14. Effect of pre-pulses of variable duration on a later pulse. Current recovery during the pulse is apparent at the higher temperature.

II. Current Recovery During the Pulse

In addition to the decay phenomena which take place only during the first few microseconds of a pulse, a slow current rise occurs during long pulses which in some cases results in a current at 50μs equal to the initial current. This recovery effect differs from that discussed earlier in that it takes place during, rather than following, the pulse. Unlike the decay effect which diminishes at high temperatures, this recovery effect decreases and becomes nonexistent at low temperatures. The general nature of this recovery and its temperature dependence are seen in Fig. 14. At 1065°K, following a small initial decay, the current recovers slowly to a value near the initial current. At 910°K this recovery is only slight and at 870°K it is not observable. This figure represents a composite of six superimposed oscilloscope traces. In obtaining these, the second pulse was of 15μs duration and fixed in position. The first pulse beginning 46μs in advance of the start of the second pulse was variable in duration and set at values of 0, 6, 16, 26, 37, and 46μs. This figure shows then the effect of a pre-pulse of variable duration on the current available during a later pulse. With no pre-pulse, the second pulse shows, of course, the same decay as is found at the beginning of the pre-pulse, when it is used. With a 6μs pre-pulse and a delay of 40μs before the second pulse, the decay in the latter pulse is diminished. This influence of the pre-pulse extending for 40μs cannot be caused by an interface voltage which remains since, as seen in Fig. 8 for this tube, the recovery after the pulse is complete in about 5μs. This delayed effect of the pre-pulse on the second pulse is not related to the decay and its subsequent recovery after the pulse, but to the current recovery which took place *during* the 6μs pre-pulse. It is essentially to the current level at the end of the pre-pulse to which the current in the second pulse decays before the second recovery begins. An examination of the 16, 26, 37, and 46μs pre-pulses, and their subsequent effect of progressively diminishing the second pulse decay, confirms this observation. Similarly, at lower temperatures where the recovery during the pre-pulse is less pronounced the effect on the second pulse is reduced. This long persisting effect of recovery during the pre-pulse has led to the hypothesis that this recovery arises due to a heating effect rather than its being associated with an electronic phenomenon. It is postulated that Joule heating occurs due to current flow

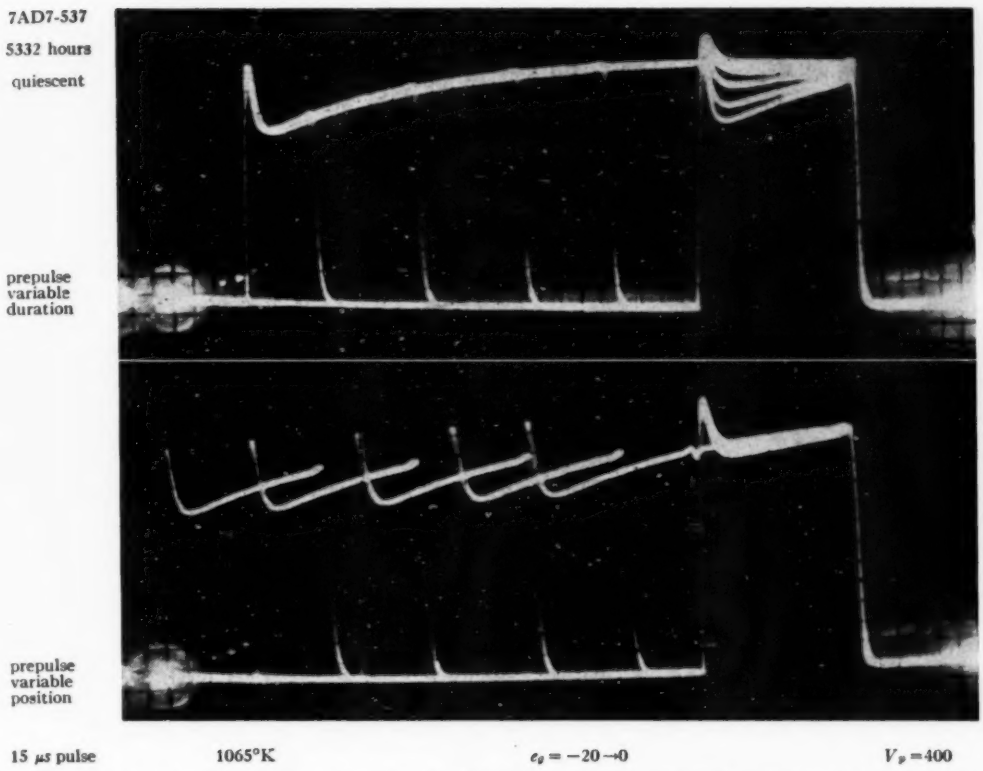


FIG. 15. Comparison of the effect of pre-pulses of variable duration and pre-pulses of variable position on the shape of a later pulse.

during the pulse; the ensuing temperature rise decreases the interface resistivity and thus also the interface voltage. Because of the thermal heat capacity of the interface, and heat loss to the adjacent base metal and coating, the interface temperature should rise slowly during the pulse and fall even more slowly following the pulse. After the initial decay the Joule heating is,

$$W = i_s^2 R_1, \tag{18}$$

and making use of Eq. (9),

$$W = \frac{e_g^2 g_m^2}{(1/R_1) + g_m + g_m R_1}. \tag{19}$$

Since g_m is always much greater than unity, the dependence of W on R_1 is determined by the third term in the denominator. Thus it is apparent that the Joule heating should diminish at lower temperatures where the interface resistivity is high. This accounts for the absence of the recovery during the pulse at low cathode temperatures.

To further illustrate the effect of a pre-pulse on the subsequent second pulse, Fig. 15 shows a comparison of the second pulse shape as affected by pre-pulses of variable duration and by pre-pulses of variable position. The upper diagram is identical with the one of Fig. 14, showing pre-pulses of 0, 6, 16, 26, 37, and 46μs duration. The lower diagram shows a second 15μs pulse, fixed in position, and a 15μs pre-pulse in five positions such that delays of 40, 30, 20, 9, and 0μs occur between the end of the pre-pulse and beginning of the second pulse. The shape of the second pulse in the latter diagram is relatively unaffected by the length of delay in the 0–40μs time range. This is in contrast to the marked effect of the duration of the prepulse as seen before. In each case the second pulse decays only to about the current level existing at the end of the pre-pulse. These observations are consistent with the Joule heating hypothesis for the current recovery during the pulse.

CONCLUSIONS

The phenomena of current decay during the pulse found in certain pentodes has been interpreted in terms of a leaky condenser interface model. That the interface consists of a homogeneous semiconductor, rather than being a region of poor coating-base metal contact, is supported by x-ray diffraction studies showing the presence of Ba_2SiO_4 in cathodes exhibiting decay. It is believed that this interface may have various degrees of activation depending on the conditions of its formation. Recovery from the current decay takes place following the end of the pulse. Measurements of the decay time constant and recovery time constant are of value in determining the interface resistance and capacitance. Semiquantitative agreement is found between these measured values and those calculated on the basis of

known properties of Ba_2SiO_4. The interface resistance thus measured is found to have the temperature dependence of a semiconductor. A recovery of current during the pulse is found when long pulses are used. Thus, because of its long time constant, its effect on subsequent pulses and its temperature dependence is believed to result from Joule heating of the interface.

In conclusion, I wish to acknowledge the generous support of this work by the ONR and by the University Research Council.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

Some Measurements of Thermal Transpiration

S. CHU LIANG
National Research Council, Ottawa, Canada
(Received July 7, 1950)

Measurements have been made of the pressure ratio, R, due to thermal transpiration in hydrogen, helium, argon, and nitrogen between a warm temperature of 297°K and cold temperatures of 77.3 or 195°K. It is found that R depends on X, the product of the pressure and tube diameter according to the empirical relation,

$$R=P_1/P_2=(AX^2+BX+R_m)/(AX^2+BX+1),$$

where A and B are constants which depend on the gas and on the warm and cold temperatures, and $R_m=(T_1/T_2)^{\frac{1}{2}}$. The R against P_2, and therefore X, curves for different gases are found to be related by a "pressure shifting factor" f.

The use of these results to correct measurements of adsorption equilibrium and vapor pressures at low pressures and low temperatures is discussed.

Vapor pressures at 77.3°±0.5°K corrected for the thermomolecular pressure effect have been measured for krypton and xenon:

Krypton	1.754±0.002 mm;
Xenon	1.93 ±0.01×10⁻³ mm.

INTRODUCTION

WHEN an investigation involving low temperature and low pressure measurement is conducted, the pressure gauge is in general operated at room temperature. The pressure of the system being investigated is considered as the same as that of the gauge after the pressure equilibrium is attained. Actually, however, this is not necessarily the case. When the pressure is low, the mean free path of the gas molecules becomes comparable with that of the connecting tubing diameter, and thermal transpiration plays an important role. It is known that, when the mean free path $\lambda \ll d$, the diameter of the tube, the pressures at both ends are practically the same; and when $\lambda \gg d$, the relation $P_1/P_2=(T_1/T_2)^{\frac{1}{2}}$ holds, where P_1 is the pressure of the part of system at T_1, and P_2 that at T_2. Thus, if a pressure gauge is at T_2, the room temperature, e.g., 25°C, and the system under investigation is at T_1, the liquid nitrogen temperature 77.3°K, then $P_1/P_2=0.51$. Obviously, if the pressure measurement is of any importance, correction must be made for this effect.

The phenomenon of transpiration had been noticed as early as before the turn of the century by Reynold).[1] It was later investigated by Knudsen, Keesom, and Weber from both the experimental and theoretical aspects; and it has been discussed in every book on the kinetic theory of gases in varying detail. However, the behavior of a gas in the intermediate pressure range when $1>P_1/P_2>(T_1/T_2)^{\frac{1}{2}}$, is not yet well understood. The present paper, therefore, wishes to point out the importance of thermal transpiration, and also to present some empirical relationships between the pressure and the size of the tube, and also among different gases.

Before presenting the experimental part, it should be noted that Weber and Schmidt[2] have suggested a semi-empirical formula for helium:

$$\begin{aligned}\log(P_1/P_2)=&\tfrac{1}{2}\log(T_1/T_2)\\&+0.18131\log[(y_1+0.1878)/(y_2+0.1878)]\\&+0.41284\log[(y_1+1.8311)/(y_2+1.8311)]\\&-0.15823\log[(y_1+4.9930)/(y_2+4.9930)],\end{aligned}\quad(1)$$

where $y=R/\lambda$, with R being the radius of the tube and λ the mean free path. λ is both pressure and temperature dependent:

$$p\lambda=(p\lambda)_0(T/273.1)^{1+n},\quad(2)$$

$(p\lambda)_0$ is the value of that quantity at 0°C, and n is obtained from viscosity η:

$$\eta/\eta_0=(T/T_0)^{\frac{1}{2}+n}.\quad(3)$$

Equation (1) is obviously not easy to use. Its accuracy for other gases according to van Itterbeek and de Grande[3] is rather limited.

[1] O. Reynold, see Maxwell, *Scientific Papers*, (1890), Vol. 2, p. 708.

[2] S. Weber and G. Schmidt, Commun. Leiden No. 246c. Rapp et Commun. 7e Congr. intern. du Froid, p. 72 (1936).

[3] A. van Itterbeek and E. de Grande, Physica 13, 289 (1937).

EXPERIMENTAL

Methods of Investigation

Two methods of investigation are available. One is called the "absolute" method used by Tompkins and Wheeler,[4] where the absolute pressure at the lower temperature region is calculated. The other is called the "relative" method, by which the thermomolecular pressure effect of two tubes of different sizes is compared. Thus, when one tube is considerably larger than the other, the effect due to the smaller tube can be measured with high accuracy provided that the pressure is not too low.

The apparatus used in the "relative" method is shown in Fig. 1. In the higher pressure region, the pressure P_1, determined by the McLeod gauge G_1 was taken as that of the cold region, and thus P_1/P_2 gave the thermomolecular pressure effect due to the narrower tube where P_2 was the pressure measured by G_2. Thus, as the pressure was lowered, P_1/P_2 decreased from unity to a minimum and then increased again to unity when P_1 was low enough so that the thermomolecular pressure effect became the same for both the wide and the narrow tubes. In the present study, this lower limit was not reached. The narrow tube B was a 1.6-mm capillary, and the wide tube C 30 mm in inside diameter and 250 mm long. The top was connected by an 8-mm tube to G_1. To insure the upper part of the capillary being at room temperature, the distance between the cold bath level and the junction of A and C was kept at least 85 mm. A glass-wool plug F was used to cover the mouth of the Dewar flask as a further precaution. Both G_1 and G_2 were calibrated with mercury before assembly. This relative method was used to measure the thermal transpiration in nitrogen, argon, hydrogen, and helium, with cold end at 195 or 77.3°K and warm end at 297±1°K.

The "absolute" method was described in great detail by Tompkins and Wheeler and involves the following concepts. It is assumed that the system can be divided into two parts, one at room temperature T_R and the other at a lower temperature T_L, the corresponding volumes being V_R and V_L, respectively. If a known quantity of gas, n moles, is introduced, then assuming that the perfect gas law holds,

$$n=(P_RV_R/RT_R)+(P_LV_L/RT_L),$$

where P_R and P_L are the pressures of the parts of the system at room and lower temperatures, respectively. P_R is measured directly by a pressure gauge, a McLeod gauge in our case, and P_L can be calculated.

The usefulness of the "absolute" method is limited by the many factors discussed by the original authors. In particular, it cannot deal with gases which are adsorbed in the pressure range to be studied. Thus, the present study with this method was limited to hydrogen and helium, with the cold end at dry ice-acetone (195°K) or liquid nitrogen (77.3°K) temperatures. The warm was again at 297±1°K. The tubing size used in the "absolute" method was 6.9 mm inside diameter.

Purification of Gases

Tank helium was passed slowly through a copper furnace maintained at 500°C to remove traces of oxygen and then through tubes containing calcium chloride, ascarite, and phosphorous pentoxide in succession to remove water, and finally through an activated charcoal trap kept at liquid nitrogen temperature to remove other impurities such as hydrocarbon vapors, nitrogen, and other noble gases. The charcoal trap was degassed at 500°C until a pressure of less than 10^{-4} mm Hg was maintained at 500°C prior to the use.

Hydrogen, nitrogen, and argon were withdrawn from tank supply and purified in the same manner as helium, but with the omission of the charcoal trap.

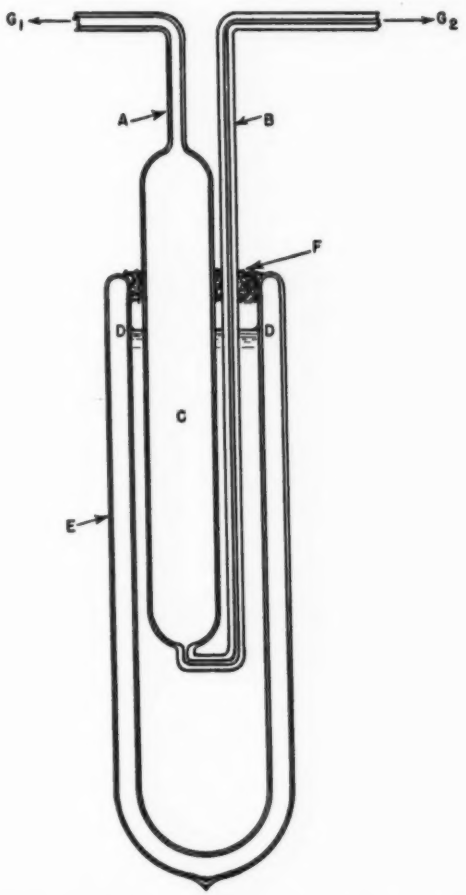


FIG. 1. Apparatus for the relative method: A—8-mm tube, B—1.6-mm capillary, C—30-mm inside diameter tube, DD—cold bath level, E—Dewar flask, F—glass-wool plug, G_1, G_2—McLeod gauges 1 and 2.

[4] F. C. Tompkins and D. E. Wheeler, Trans. Faraday Soc. 29, 1248 (1933).

TABLE I. Thermomolecular pressure effect in hydrogen, $d=6.9$ mm (absolute method).

195°K			77.3°K		
P_2 mm Hg	P_1/P_2	Experimental error (%)	P_2 mm Hg	P_1/P_2	Experimental error (%)
0.141	1.002	2	0.28	1.024	5
0.111	0.980	2	0.19	1.007	5
0.0926	0.968	2	0.094	1.002	5
0.0714	0.939	3	0.067	0.916	5
0.0567	0.923	3	0.0487	0.858	5
0.0475	0.905	3	0.0258	0.781	7
0.0328	0.875	4	0.0189	0.764	7
0.0246	0.865	4	0.0141	0.678	7
0.0170	0.842	5	0.00914	0.616	10
0.00996	0.824	5	0.00421	0.538	10

TABLE II. Thermomolecular pressure effect in helium, $d=6.9$ mm (absolute method).

195°K			77.3°K		
P_2 mm Hg	P_1/P_2	Experimental error (%)	P_2 mm Hg	P_1/P_2	Experimental error (%)
0.424	1.00	2	0.292	1.01	5
0.260	1.00	2	0.196	0.975	5
0.214	0.986	2	0.149	0.956	5
0.196	0.974	2	0.0922	0.882	5
0.0965	0.927	2	0.0716	0.817	5
0.0678	0.904	3	0.0488	0.802	5
0.0496	0.890	3	0.0338	0.731	7
0.0353	0.848	4	0.0268	0.679	7
0.0258	0.818	4	0.0137	0.628	7
0.0185	0.81	5	0.00834	0.512	10

RESULTS AND INTERPRETATION

I. The Absolute Method

With this method, we encountered the same difficulties as Tompkins and Wheeler, namely, to keep a relatively sharp temperature boundary and to obtain a sufficiently high accuracy. Our measurements for hydrogen are given in Table I and those for helium in Table II. The error involved in individual readings is also given.

II. The Relative Method

In the pressure region where the thermomolecular pressure effect in the 30-mm tube is negligible, the value $R(=P_1/P_2)$ can be determined without ambiguity. When the pressure P_2 is below 0.05 mm, correction must be made for both P_1 and R. This correction has been based upon an empirical estimate which has been found to be satisfactory and is discussed in the later part of this paper. However, the introduction of this correction will necessarily increase the experimental error. Thus, for $P_2<0.05$ mm the values of R are not very reliable.

In Figs. 2 and 3, R is plotted against P_2, for different gases at 77.3 and 195°K. R is presented as a function of P_2 rather than of P_1 because P_2 is the pressure which would ordinarily be read from a pressure gauge, in the present case, a McLeod gauge. These curves may thus serve as the calibration curves.

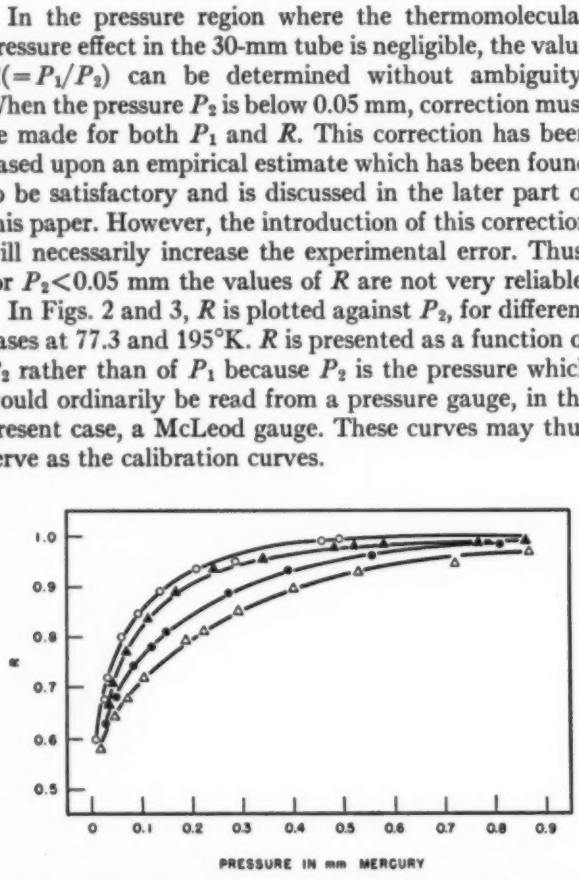


FIG. 2. Thermomolecular pressure ratio at 77.3 *vs* 297°K, $d=1.6$ mm. Open circle—nitrogen, closed circle—hydrogen, closed triangle—argon, open triangle—helium.

It is of importance to notice that each of the four gases gives a separate curve, and the differences are large enough to allow one to characterize the gases.

The usefulness of these curves has been very much limited because the value of R is a function not only of a particular gas, but also of temperature, pressure, tube diameter, and probably of tube material and state of wall surface (commercial Pyrex glass tubes are used in the present research). The formula proposed by Weber [Eq. (1)] is useful only when dealing with helium. In this paper, the pressure-tube diameter $(p-d)$ dependence and the relation between different gases will be discussed.

Pressure and Tube Diameter

From the theoretical aspect it is well known that this effect becomes important only when the mean free path of the gas molecules is appreciable when compared with the tube diameter. It is thus reasonable to assume that for a particular gas subject to a particular temperature difference, the value of R will be the same when the ratio λ/d has the same value, notwithstanding the absolute values of λ and d. The functional relation between R and λ or d may, however, still be complex. It is widely known that in the rarified state the relationship,

$$p\lambda=\text{constant}, \tag{4}$$

holds true for a certain gas at a particular temperature.

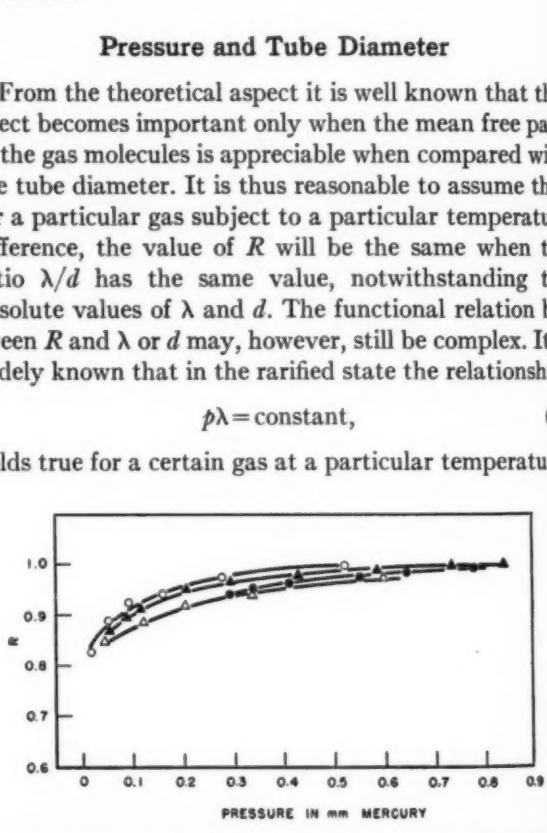


FIG. 3. Thermomolecular pressure ratio at 195 *vs* 297°K, $d=1.6$ mm. Open circle—nitrogen, closed circle—hydrogen, closed triangle—argon, open triangle—helium.

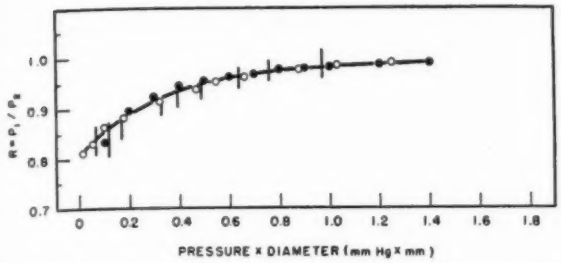


FIG. 4. Thermomolecular pressure ratio in hydrogen at 195 *vs* 297°K. Closed circle: (Tompkins and Wheeler's data) $d=10$ mm. Open circle: $d=1.6$ mm (relative method). Line $d=6.9$ mm (absolute method).

Thus, when

$$\lambda_1/d_1=\lambda_2/d_2, \tag{5}$$

it follows:

$$p_1d_1=p_2d_2. \tag{6}$$

In other words, for a certain temperature difference, e.g., 77.3 *vs* 298°K, when R is plotted against the product of pressure and tube diameter, i.e., $p\cdot d$, each gas will give only one curve, regardless of the absolute value of d. This reasoning has been found to be true and two illustrations are given in Figs. 4 and 5, where R is plotted against $p\cdot d$, with p in mm of mercury and d in mm.

Figure 4 gives the data for 195°K and Fig. 5 for 90°K. Both are for hydrogen. Figure 4 contains three sets of data. Two of these are based on our own measurements, one with 6.9-mm tube by the absolute method and the other with 1.6-mm tube by the relative method (tubing diameter measured by mercury filling), and the third set is a replot of Tompkins and Wheeler's data, which involves a tube diameter of 10 mm. In Fig. 5, two sets are the measurements of van Itterbeek and de Grande[3] using $d=1.236$ and 0.860 mm, respectively, and the third set is due to Tompkins and Wheeler, again using $d=10$ mm. In plotting Tompkins and Wheeler's data, the claimed reproducibility of individual readings was noticed. However, it was also noticed that the value of $R=1.034$ at 90°K and $R=1.004$ at 193°K were given by the authors. These values are obviously wrong, since they cannot exceed unity. It was concluded that a systematic error must be involved and thus 5 percent for the 90°K and 1 percent for the 193°K data were arbitrarily chosen. Their data were corrected by dividing the

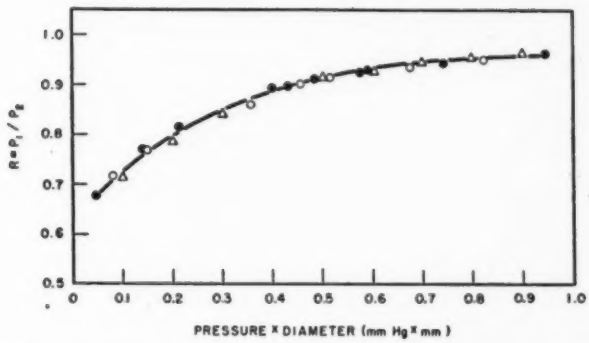


FIG. 5. Thermomolecular pressure ratio in hydrogen at 90 *vs* 297°K. Triangle: (Tompkins and Wheeler's data) $d=10$ mm. Circles: (van Itterbeek and de Grande's data) open—$d=0.860$ mm, closed —$d=1.236$ mm.

90°K R value by 1.05 and the 193°K value by 1.01 accordingly. The good agreement is shown in both cases.

Having established this relationship for pd against R, it becomes possible to correct the lower pressure measurements ($P_2<0.05$ mm) of the present work for the effect in the wide tube side. This however introduces an experimental error of some 1–2 percent. In Tables III to VI are shown some of the experimental results obtained by the relative method. The column heading "estimated correction factor" is the factor estimated by this relationship.

Correlation between Different Gases

An attempt has been made to correlate the thermomolecular pressure effect for the various gases investigated, but at the present time this correlation is not complete. At either 77.3 or 195°K,[5] the four curves can be reduced to a single curve by changing the pressure scale. For instance, the nitrogen curve can be made to coincide with that of helium if the nitrogen pressure scale is multiplied by 2.12 at 195°K or 2.80 at 77.3°K. These "pressure shifting factors" are given in Table VII.

Each of the R *vs* $p\cdot d$ plot is found to be expressed satisfactorily by the empirical formula

$$R=(AX^2+BX+R_m)/(AX^2+BX+1), \tag{7}$$

TABLE III. Thermomolecular pressure ratio in nitrogen, $d=1.6$ mm (relative method).

77.3°K				195°K			
P_2 mm	P_1/P_2 Apparent	Estimated correction factor	R ($=P_1/P_2$) Corrected	P_2 mm	P_1/P_2 Apparent	Estimated correction factor	R ($=P_1/P_2$) Corrected
0.4921	0.993	1.00	0.993	0.2790	0.973	1.00	0.973
0.2569	0.959	1.00	0.959	0.0904	0.923	1.00	0.923
0.1349	0.891	1.00	0.891	0.0537	0.888	1.00	0.888
0.0322	0.735	0.981	0.721	0.0171	0.863	0.960	0.829
0.0255	0.705	0.960	0.677	0.0101	0.863	0.932	0.804
0.00636	0.709	0.839	0.595				

TABLE IV. Thermomolecular pressure ratio in argon, $d=1.6$ mm (relative method).

77.3°K				195°K			
P_2 mm	P_1/P_2 Apparent	Estimated correction factor	R ($=P_1/P_2$) Corrected	P_2 mm	P_1/P_2 Apparent	Estimated correction factor	R ($=P_1/P_2$) Corrected
0.5798	0.984	1.00	0.984	0.4274	0.978	1.00	0.978
0.4826	0.979	1.00	0.979	0.2944	0.963	1.00	0.963
0.3395	0.951	1.00	0.951	0.2074	0.952	1.00	0.952
0.1626	0.889	1.00	0.889	0.1155	0.911	1.00	0.911
0.0729	0.780	1.00	0.780	0.0862	0.895	1.00	0.895
0.0354	0.689	0.971	0.669	0.0600	0.868	1.00	0.868

[5] The cold end temperature will be used to indicate the set of data being discussed because the warm end has been at a constant temperature of 297±1°K.

TABLE V. Thermomolecular pressure ratio in hydrogen, $d=1.6$ mm (relative method).

77.3°K P_2 mm	77.3°K P_1/P_2 Apparent	77.3°K Estimated correction factor	77.3°K R $(=P_1/P_2)$ Corrected	195°K P_2 mm	195°K P_1/P_2 Apparent	195°K Estimated correction factor	195°K R $(=P_1/P_2)$ Corrected
0.8089	0.987	1.00	0.987	0.6462	0.983	1.00	0.983
0.5570	0.962	1.00	0.962	0.4137	0.962	1.00	0.962
0.3913	0.931	1.00	0.931	0.3402	0.952	1.00	0.952
0.2718	0.885	1.00	0.885	0.2954	0.939	1.00	0.939
0.0837	0.750	0.989	0.742	0.1109	0.881	1.00	0.881
0.0501	0.707	0.965	0.683	0.0646	0.872	0.989	0.862

TABLE VI. Thermomolecular pressure ratio in helium, $d=1.6$ mm (relative method).

77.3°K P_2 mm	77.3°K P_1/P_2 Apparent	77.3°K Estimated correction factor	77.3°K R $(=P_1/P_2)$ Corrected	195°K P_2 mm	195°K P_1/P_2 Apparent	195°K Estimated correction factor	195°K R $(=P_1/P_2)$ Corrected
0.8683	0.968	1.00	0.968	0.5997	0.974	1.00	0.974
0.4025	0.901	1.00	0.901	0.3386	0.939	1.00	0.939
0.2909	0.850	1.00	0.850	0.2049	0.918	1.00	0.918
0.1899	0.795	1.00	0.795	0.1212	0.889	1.00	0.889
0.1047	0.740	0.975	0.721	0.0425	0.888	0.965	0.847
0.0469	0.703	0.921	0.647	0.00898	0.936	0.88	0.82
0.0185	0.728	0.800	0.582				

where $X=p\cdot d$, $R_m=(T_1/T_2)^{\frac{1}{2}}$, the minimum value of R, and A and B are functions of gases and temperature. When the pressure is high, $X\to\infty$, $R\to 1$; i.e., the pressure at the hot and cold regions are practically the same, as generally observed. When the pressure is low, $X\to 0$, $R\to R_m$, again the situation to be expected. It must be pointed out in this connection, that Eq. (7) is, at best, only an approximate expression. It is obvious that when $\lambda\gg d$, $R=R_m$. Therefore, X should be replaced by $(X-x)$, where x corresponds to the pressure at which $\infty>\lambda\gg d$ and $P_1/P_2=1$. We have not been able to determine the value of x accurately. We have found, however, that without the correction term x, Eq. (7) is applicable at 195°K for $R\geq 0.85$, where $R_m=0.81$, and at 77.3°K for $R\geq 0.7$, where $R_m=0.51$.

Using nitrogen as the standard and extracting the "shifting factor" f gives Eq. (7) in the form

$$R=P_1/P_2=[\alpha(X/f)^2+\beta(X/f)+R_m]/[\alpha(X/f)^2+\beta(X/f)+1], \quad (8)$$

where α and β are constants independent of the gas investigated, and f is a function of the gas and temperature difference. Taking f equal to unity for nitrogen at both 77.3 and 195°K, the values of α and β are given in Table VIII.

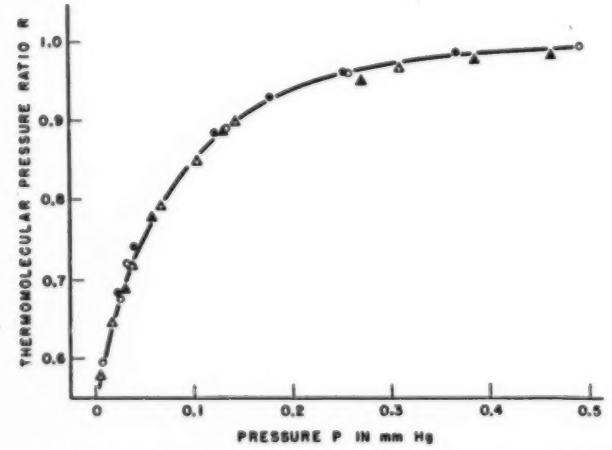


FIG. 6. "Reduced plot" of thermomolecular pressure ratio at 77.3 *vs* 297°K, $d=1.6$ mm, with nitrogen curve as standard. Open circle—nitrogen, closed circle—hydrogen. Closed triangle—argon, open triangle—helium.

Figures 6 and 7 are the reduced* plots of Figs. 2 and 3. A good agreement is found for the various gases in both cases.

The shifting factor f is undoubtedly related to the collision diameter of a gas molecule although this relationship has not been definitely established.

APPLICATION

The thermomolecular pressure ratio is apparent in all pressure measurements where the measuring instrument differs in temperature from the vessel in which the pressure is required. This effect is of importance in many fields, and two may be mentioned.

TABLE VII. "Pressure shifting factor" f with nitrogen as the standard.

Gas		Nitrogen	Argon	Hydrogen	Helium
f values	77.3°K	1	1.25	2.2	2.80
	195°K	1	1.30	1.8	2.12

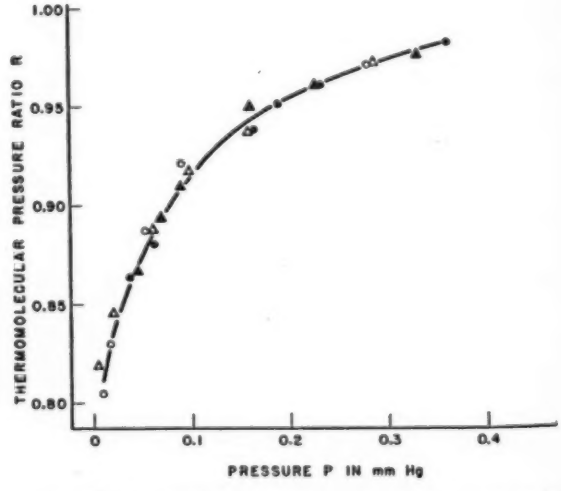


FIG. 7. "Reduced plot" of thermomolecular pressure ratio at 195 *vs* 297°K, $d=1.6$ mm, with nitrogen curve as standard. Open circle—nitrogen, closed circle—hydrogen, closed triangle—argon, open triangle—helium.

* The term "reduced plot" is used because many curves are reduced to a single curve.

TABLE VIII. Empirical constants α and β.

Temperature	α	β
195°K	22.76	5.00
77.3°K	29.5	12.5

I. Adsorption

The effects of this phenomenon, discussed in some detail by Tompkins and Wheeler, include the interpretation of kinetics and the calculation of Clapeyron-Clausius heat of adsorption. It must be pointed out that without this correction, the large volume of accurately determined low pressure adsorption measurements will be meaningless. Thus, the adsorption study of ethylene at liquid air temperatures on various catalysts and the adsorption of gases on evaporated metal films, if not corrected, may all need new interpretation. In the literature, correction for this effect has seldom been mentioned.

II. Vapor Pressure Measurements

To illustrate how the thermal transpiration may affect the vapor pressure measurement, the vapor pressure of two gases has been measured at 77.3±0.5°K, namely, krypton with a reasonably high vapor pressure, and xenon with a lower vapor pressure. Both gases used were Linde Air products, refined grade, of 99.95 percent purity.

With krypton, no further purification was made. The measurement was conventional. A value 1.754±0.002 mm was obtained. A comparison with the values reported by other investigators is given in Table IX.

The xenon was purified further by repeatedly solidifying and pumping off the noncondensable gas. By varying the diameter of the lead-in tube from 1.6 mm to 30 mm, the vapor pressure values were recorded as 3.79 $\times 10^{-3}$ and 2.21$\times 10^{-3}$ mm, all with excellent reproducibility. The thermal transpiration effect in xenon was then determined and appropriate correction was made for each reading. A unique value 1.93±0.01$\times 10^{-3}$ mm was obtained in each case. No experimental value at this temperature has been reported in the literature. The present results are shown in Table X. It is evident, therefore, that every value between 1.9 and 3.8$\times 10^{-3}$ mm could have been reported as an accurate result, and yet been in error without proper correction for thermal transpiration.

TABLE IX. Vapor pressure of krypton at 77.3°K.

Author	Peters and Weil[a]	Keesom, Mazur, and Meiheizen[b]	Justi[c]	Liang
Vapor pressure in mm	1.90±0.01	1.90±0.01	2.26±0.02	1.754±0.002
Remarks	Extra-polation	Extra-polation	Inter-polation	Direct measure-ment

[a] K. Peters and K. Weil, Z. physik. Chem. **A148**, 27 (1930).
[b] Keesom, Mazur, and Meiheizen, Physica **2**, 669 (1935).
[c] E. Justi, Physik. Z. **36**, 571 (1935).

TABLE X. Measurement of xenon vapor pressure at 77.3°K.[a]

Leading tube diameter mm	Apparent vapor pressure P' mm Hg	Thermomolecular pressure ratio R	Corrected vapor pressure $P(=P'\times R)$ mm Hg
1.6	3.79$\times 10^{-3}$	0.510	1.93$\times 10^{-3}$
30	2.21	0.88	1.94

[a] The pressure measurements are accurate to 2 percent.

ACKNOWLEDGMENT

The author is grateful to Dr. A. Cambron for his continuous interest and encouragement and to Dr. G. L. Kington for some interesting discussions and his help in preparing this paper.

On the Directional Patterns of Polystyrene Rod Antennas*

R. B. WATSON
Defense Research Laboratory, The University of Texas, Austin, Texas
(Received June 13, 1950)

Experimental and theoretical directional patterns are compared for three polystyrene rod antennas of length variable from about 4 to 10 wavelengths, retaining relative shape in the length direction. The comparison shows that the apparent index of refraction, n, decreases with increasing length. Accordingly, the directional patterns are less sharp than might be expected from the estimated values of n which are constant for a constant relative shape.

THE use of polystyrene rods as antennas has been considered both experimentally and theoretically in earlier articles.[1,2] Experimental measurements are here described for three specific forms, and the data are compared with theoretical calculations. These three antennas are illustrated by Fig. 1. The exposed length of the antenna in each case was at first 12 inches, as shown by Fig. 1, and was then reduced by 1-inch increments until it reached a minimum of 6 inches. In this reduction of length, the relative proportions between the tapered part and the uniform part were retained, so that always one-half of the exposed length was tapered for Poly Form 12A, one-third for Poly Form 12B, and two-thirds for Poly Form 12C. The remaining portion in each case was kept at one-half the maximum thickness of the form. Each form fitted closely into a standard ½-in.×1-in. wave guide, and was there terminated in a three-inch taper.

Experimental data in both E- and H-planes were obtained from patterns for these three forms for the half-width of the major lobe at the −10-db point, angular positions of minor lobes, and angular positions of minima, for a wavelength of 3.2 cm.

It was evident from comparison with earlier curves,[1] that the experimental data corresponded to computed values obtained by using an apparent index of refraction less than $n=1.10$. Calculations were made for $n=1.05$ (0.01) 1.10. The positions of the minor lobes, which were obtained to a good approximation as the maxima of the integral I_2,[2] are shown in Fig. 2 *versus* length of the antenna. Figure 3 shows the several calculated minima as determined by the minima of the same integral I_2. Figure 4 shows the calculated half-widths of the major lobe at the −10-db point for the E- and H-planes. In the E-plane three values of $ka/2$ are required for the three different forms, but detailed calculations showed these values to differ only slightly; and hence only one set of curves is drawn. Figure 5 shows the theoretical curves for the height of the first minor lobe *versus* the length of the antenna for the E- and H-planes. Again, it was found necessary to plot only one set of curves for the three values of $ka/2$.

The calculated values are compared with the data for the Poly Form 12A in Figs. 6–9. In the first three of these figures theoretical curves are included for the expected value of n only, 1.08. Comparisons of Figs. 2 and 4 with Fig. 6, and Fig. 3 with Fig. 7, show that the

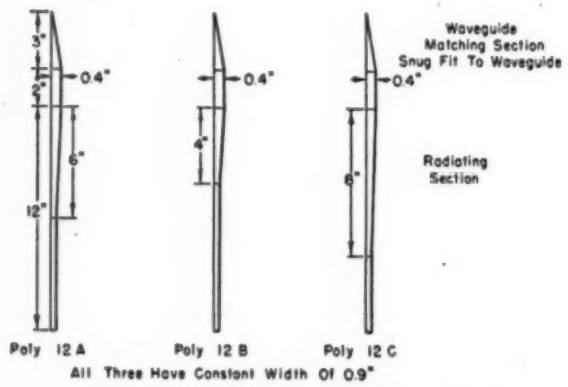


FIG. 1. Dimensions for polystyrene rod antennas.

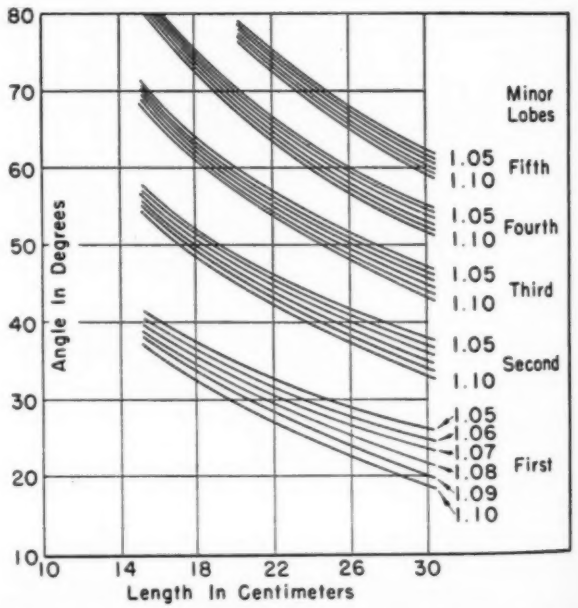


FIG. 2. Computed positions of minor lobes relative to major lobe *versus* antenna length, with apparent index of refraction as parameter.

* The work described in this paper was done at Defense Research Laboratory, under the sponsorship of the Bureau of Ordnance, Navy Department, Contract NOrd-9195.

[1] R. B. Watson and C. W. Horton, J. Appl. Phys. **19**, 661 (1948).

[2] R. B. Watson and C. W. Horton, J. Appl. Phys. **19**, 836 (1948).

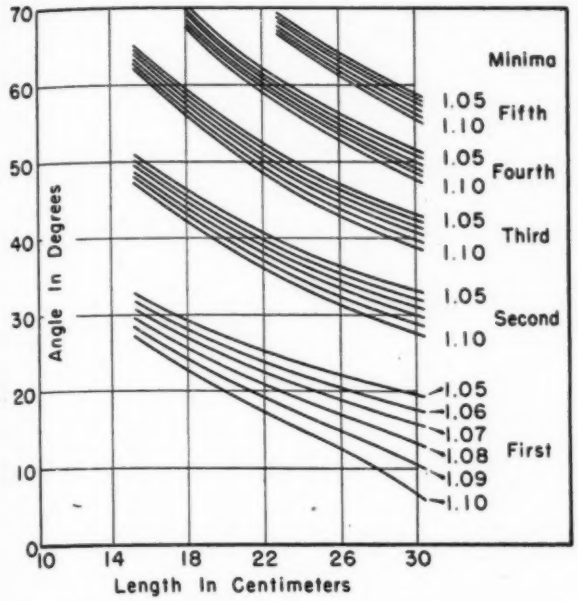


FIG. 3. Computed positions of minima relative to major lobe *versus* antenna length, with apparent index of refraction as parameter.

data vary considerably from the computed values for $n=1.08$, and indeed from a constant value of n, the trend being from relatively large values of n at short lengths to smaller values of n at long lengths. The same trend is apparent in Fig. 8 despite the scatter of data, in that the straight lines fitted to the data cross the set of theoretical curves in the direction of decreasing n.

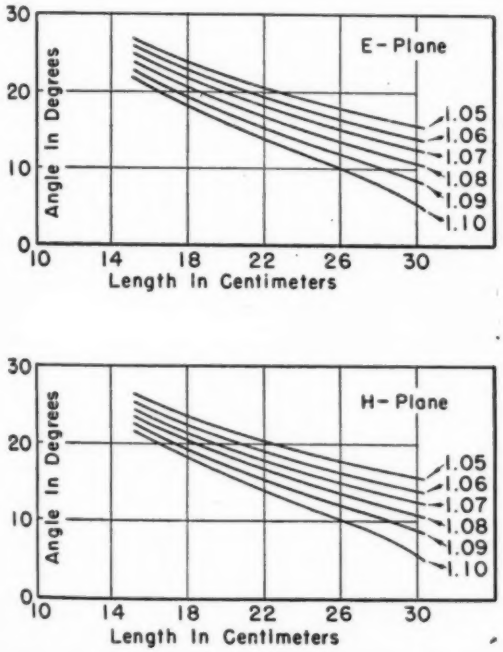


FIG. 4. Computed half-width of major lobe, 10 db below maximum, *versus* antenna length, with apparent index of refraction as parameter.

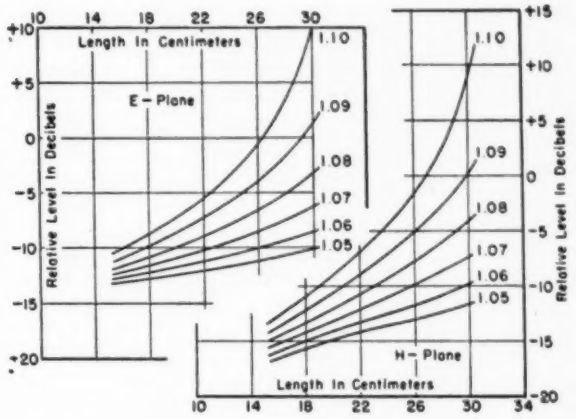


FIG. 5. Computed height of first minor lobe relative to major lobe *versus* antenna length, with apparent index of refraction as parameter.

The values of n to be expected for these antennas were obtained by examination of Fig. 5 in the paper by G. Wilkes.[3] The values of a/λ are 0.198 for Poly Form 12A, 0.185 for Poly Form 12B, and 0.212 for Poly Form 12C. The expected values of n are respectively 1.08, 1.07, and 1.09, independent of the length. As a measure of deviation from these expected values, values of n were chosen from the calculated curves to match each experimental point, for all three poly forms. These values are plotted in Fig. 9, for the E-plane only, for the half-width of

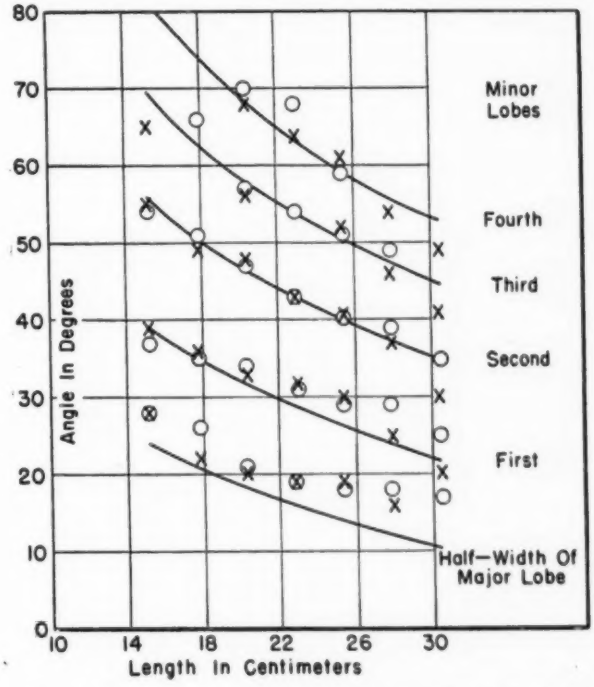


FIG. 6. Positions of minor lobes and half-width of major lobe at −10-db point *versus* antenna length for Poly 12A. ○—experiment, E-plane, ×—experiment, H-plane. —— computed values, $n=1.08$.

[3] G. Wilkes, Proc. Inst. Radio Engrs. 36 207 (1948).

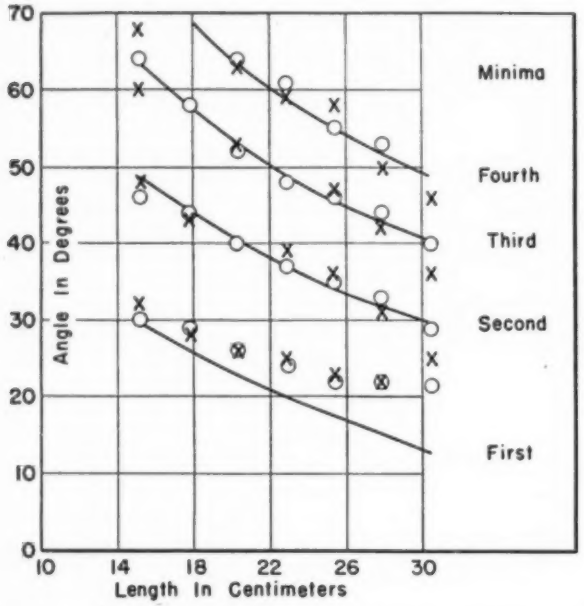


FIG. 7. Positions of minima of antenna patterns *versus* antenna length, for Poly 12A. ○—experiment, *E*-plane. ×—experiment, *H*-plane. —— computed values, $n=1.08$.

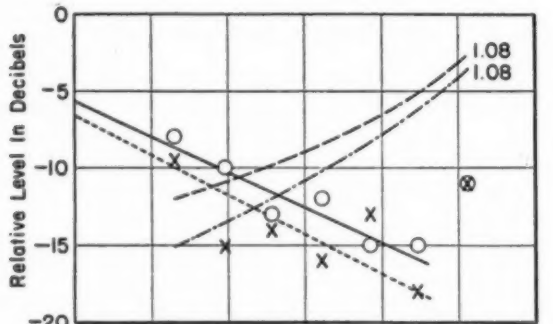


FIG. 8. Height of first minor lobe relative to major lobe *versus* antenna length, for Poly 12A. ○,—— experiment, *E*-plane. — — — — theory, $n=1.08$, *E*-plane. ×,- - - - experiment, *H*-plane. — — — — theory, $n=1.08$, *H*-plane.

the major lobe at the −10-db point, the position of the first minimum, and the position of the first minor lobe. The trend from larger to smaller values of n is once

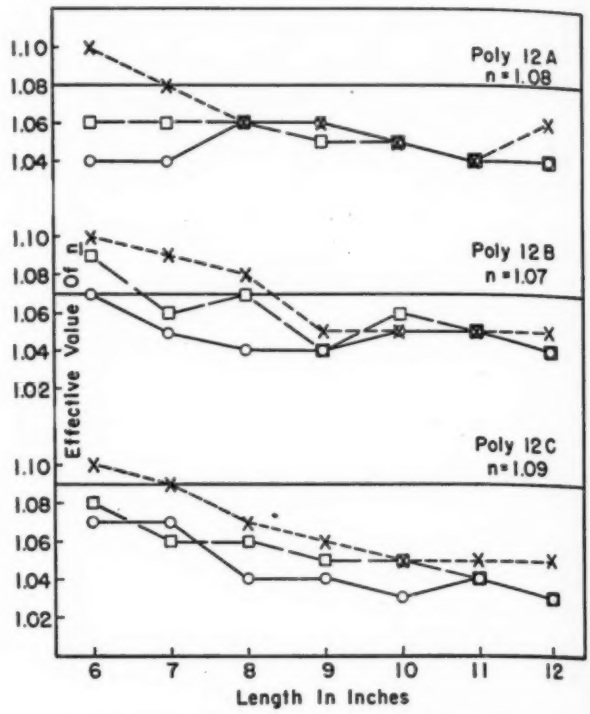


FIG. 9. Effective values of index of refraction *versus* antenna length; comparison of experiment and theory. ○—half-width (at −10 db) of major lobe. ×—position of first minor lobe. □—position of first minimum.

again evident; the data cluster about the expected values only at the shorter lengths. Estimated values for data lying outside the range of the theoretical curves were always in the direction of smaller values of n. Similar data for the *H*-plane show the same trend.

It was pointed out earlier[1] that agreement between theory and experiment was not too good for these antennas for lengths of the order of 10 wavelengths. It is concluded from this study that the increase in length of a particular tapered polyrod antenna from about 4 to 10 wavelengths, maintaining relative shape in the length direction, leads to somewhat lower values of n than indicated by theory, and to values of n which decrease with increasing length, and therefore to directional patterns which are less sharp than expected for the longer lengths.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

Asymptotic Approximation for the Normal Modes in Sound Channel Wave Propagation

N. A. HASKELL
Base Directorate for Geophysical Research, Air Force Cambridge Research Laboratories, Cambridge, Massachusetts
(Received June 23, 1950)

Asymptotic methods are used to find approximate solutions of the acoustic wave equation in a medium in which the velocity is a continuously variable function of one coordinate. It is shown that, when the velocity function has a minimum, undamped normal mode solutions exist and are closely analogous to the internally reflected waves in the case of a medium made up of discrete layers. By converting the sum of the high order normal modes into an equivalent integral, it is shown that superposition of these modes leads to geometrical ray theory modified by diffraction in a manner that may be computed from the incomplete fresnel and airy integrals.

INTRODUCTION

In a number of geophysical problems, one is concerned with cases of wave propagation in a medium in which the propagation velocity is a function of one coordinate and has one or more minima. In the atmosphere, the velocity of sound has a minimum in the lower stratosphere, goes through a maximum at an altitude of 40 to 60 kilometers, and has a second minimum at about 80 kilometers. A sound velocity minimum also exists in the ocean at a depth of about 4000 feet.[1] In the earth, the velocities of both longitudinal and transverse seismic waves in the silicate mantle have usually been assumed to increase with depth, but recent evidence suggests[2] that the "granitic" layer of the continents may actually consist of lower velocity material beneath a higher velocity layer. In the extremely heterogeneous material of the superficial layers with which one is concerned in seismic exploration, the occurrence of velocity minima and maxima is the rule, although the general trend is an increase of velocity with depth.

When the relative change in the properties of the medium is small within a wavelength, all of these cases may be formulated by wave equations of the same simple type, the difference being in the boundary conditions to be satisfied. In the present discussion, we shall use the boundary conditions appropriate to the atmospheric case. The solution for propagation in the ocean would involve only minor changes, but the seismological case is somewhat more complex owing to the coupling between longitudinal and transverse waves at the surface.

1. WAVE EQUATION

When the effect of gravity and the consequent variation of initial density is not ignored, the hydrodynamic equations of motion do not, in general, possess irrotational solutions that can be expressed in terms of a velocity potential.[3] However, if the relative change of the initial density within a wavelength is small, the motion may be approximated *locally* by a velocity potential φ satisfying the simple wave equation,

$$c^2(z)\nabla^2\phi-\partial^2\phi/\partial t^2=0, \tag{1.01}$$

with the particle velocity vector V and the pressure perturbation p given by

$$V=-\text{grad}\phi, \tag{1.02}$$

$$p=\rho(z)(\partial\phi/\partial t), \tag{1.03}$$

where $c(z)$ is the local velocity of sound and $\rho(z)$ is the local density, both functions of the altitude z. The acoustic energy flux density is

$$J=pV=-\rho(z)(\partial\varphi/\partial t)\,\text{grad}\,\varphi. \tag{1.04}$$

At a later stage, we shall compare this expression for J with that derived on the basis of ray geometry and find that they differ by a factor $\rho(h)/\rho(z)$, where $\rho(h)$ is the density at the altitude, h, of the source. Thus, although the ignored gravitational terms are small for sufficiently high frequencies, there is a cumulative effect that may become quite large, if the difference between z and h is large. The comparison with the geometrically derived energy flux suggests that, to an approximation sufficient for our present purpose, the effect of the variation of initial density may be taken into consideration by applying $\rho(h)/\rho(z)$ as a correction factor to J. Since we have, approximately,

$$|J|\simeq|p|^2/\rho(z)c(z)\simeq|V|^2\rho(z)c(z), \tag{1.05}$$

the corresponding correction factor for p and V is $[\rho(h)/\rho(z)]^{\frac{1}{2}}$.

A more rigorous consideration of the gravity-dependent terms in the equations of motion shows that for very low frequencies the velocity of propagation is dependent upon the frequency. However, the dispersion is very small if

$$|(\lambda/2\pi)d(\log\rho)/dz|\ll 1, \tag{1.06}$$

where λ is the wavelength. The average value of $d(\log\rho)/dz$ to an altitude of 50 kilometers in the atmosphere is about $-0.14\ \text{km}^{-1}$, so that Eq. (1.06) is equivalent to $\lambda\ll 45$ km, which implies that the dispersion may be ignored for frequencies very much greater than 0.0067 cycles. Since the asymptotic methods that we shall use provide a somewhat more severe low frequency limitation than this, we shall base the development on the

[1] Ewing, Worzel, and Pekeris, Geol. Soc. Am. Memoir 27 (1948).
[2] B. Gutenberg, Science 111, 29 (1950).
[3] H. Lamb, *Hydrodynamics*, sixth edition (Cambridge University Press, London, 1932), 541–556; P. G. Bergmann, J. Acoust. Soc. Am. 17, 329 (1946).

simple wave equation (1.01) with the density correction factor $[\rho(h)/\rho(z)]^{\frac{1}{4}}$ applied to p and V so that Eqs. (1.02) and (1.03) become

$$V=-[\rho(h)/\rho(z)]^{\frac{1}{4}}\,\mathrm{grad}\phi, \tag{1.07}$$

$$p=[\rho(h)\rho(z)]^{\frac{1}{4}}(\partial\phi/\partial t). \tag{1.08}$$

2. NORMAL MODE SOLUTIONS OF THE WAVE EQUATION

Periodic solutions of Eq. (1.01) in a cylindrical coordinate system with axis passing through the source may be written in the form

$$\phi(r, z, t, k)=\exp(i\omega t)J_0(kr)F(z, k), \tag{2.01}$$

where $F(z, k)$ satisfies the equation,

$$(d^2F/dz^2)+\{[\omega/c(z)]^2-k^2\}F=0. \tag{2.02}$$

The boundary conditions are:

a. At the ground surface ($z=0$) the vertical component of the particle velocity vanishes.

b. $V(z)\rightarrow 0$ as $z\rightarrow\infty$.

c. In the neighborhood of the source ($r=0, z=h$) the integrated solution,

$$\phi(r, z, t)=\int_0^\infty \phi(r, z, t, k)dk, \tag{2.03}$$

must reduce to the value appropriate to a point source, namely,

$$\phi\rightarrow A[r^2+(z-h)^2]^{-\frac{1}{4}} \times\exp i\omega\{t-[r^2+(z-h)^2]^{\frac{1}{2}}/c(h)\}. \tag{2.04}$$

Designating the regions below and above the source by subscripts (1) and (2), respectively, condition (a) is

$$(dF_1/dz)_0=0, \tag{2.05}$$

condition (b) requires that F_2 be a particular solution of Eq. (2.02) such that, with the factor $\exp(i\omega t)$, it represents an upward traveling wave at large values of z, and condition (c) is satisfied[4] if we impose the conditions at $z=h$

$$F_1(h, k)=F_2(h, k), \tag{2.06}$$

$$(dF_1/dz)_h-(dF_2/dz)_h=2Ak. \tag{2.07}$$

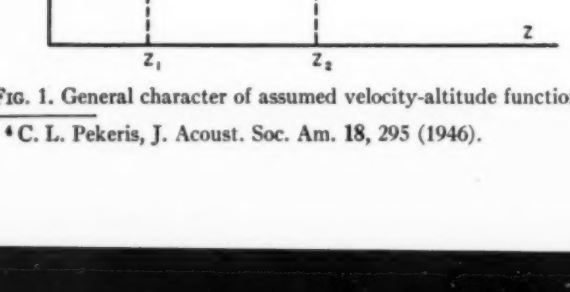


FIG. 1. General character of assumed velocity-altitude function.

[4] C. L. Pekeris, J. Acoust. Soc. Am. 18, 295 (1946).

Let $M(z, k)$ and $N(z, k)$ be two linearly independent solutions of Eq. (2.02) that behave asymptotically for large values of z (with the factor $\exp(i\omega t)$) like downward and upward traveling waves respectively. The boundary conditions are then satisfied by

$$F_1(z, k)=[2AkN(h, k)/bN'(0, k)] \times[N(z, k)M'(0, k)-M(z, k)N'(0, k)], \tag{2.08}$$

$$F_2(z, k)=[2AkN(z, k)/bN'(0, k)] \times[N(h, k)M'(0, k)-M(h, k)N'(0, k)], \tag{2.09}$$

where the primes denote differentiation with respect to z, and $b=MN'-M'N$ is a constant by virtue of the Abelian identity which holds between any two linearly independent solutions of Eq. (2.02).

It has been shown by Pekeris[1,4] that the integral (2.03) for ϕ can be reduced by replacing $J_0(kr)$ by $[H_0^{(1)}(kr)+H_0^{(2)}(kr)]/2$ followed by a transformation of the path of integration in the complex k plane, to a sum of residues taken at the zeros of $N'(0, k)$ plus certain integrals around cuts in the k plane. The contribution to ϕ of the residues, i.e., the "normal mode" solutions, is

$$\phi_{\mathrm{res}}=-2\pi iA\,\exp(i\omega t)\sum_n k_nH_0^{(2)}(k_nr) \times N(z, k_n)N(h, k_n)/N(0, k_n)[\partial N'(0, k)/\partial k]_{k_n}, \tag{2.10}$$

where the k_n's are the roots of $N'(0, k_n)=0$ and the constant b has been eliminated by use of the relation $b=-N(0, k_n)M'(0, k_n)$. Since Eq. (2.10) is unaltered by interchanging z and h, the expression is applicable both above and below the level of the source.

At large distances from the source the radial factor in Eq. (2.10) behaves like

$$H_0^{(2)}(k_nr)\rightarrow(2/\pi k_nr)^{\frac{1}{2}}\exp[-i(k_nr-\pi/4)]. \tag{2.11}$$

In conjunction with the factor $\exp(i\omega t)$ this must represent a progressive wave traveling in the direction of increasing r, so that the real part, of k_n must be positive and the imaginary part negative or zero. Those roots, if any exist, that have a finite imaginary part lead to an exponential decay factor, so that for large values of r the significant terms in Eq. (2.10) are those for which k_n is real.

3. APPROXIMATE EXPRESSIONS FOR THE FUNCTION $N(z, k)$

We now assume that the velocity *vs* altitude function has the general form shown in Fig. 1. This function and its first derivative are taken to be continuous. Above the altitude z_2 the velocity is assumed to decrease monotonically. This means that we are ignoring any refraction of wave energy back from the higher velocity levels above the second minimum. In addition to simplifying the problem, this assumption is probably justified physically, since attenuation becomes significant at very high altitudes and is believed to prevent any measurable return at frequencies high enough to insure the validity of the asymptotic methods we shall use.

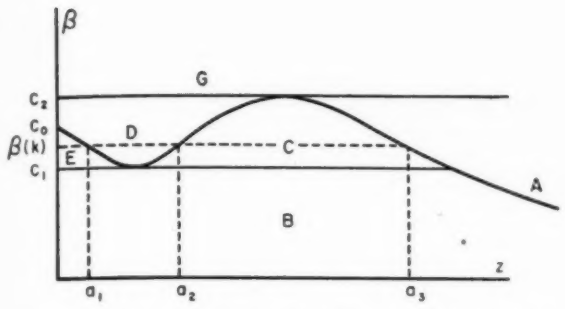


FIG. 2. Division of (β, z) plane into domains defined by the relative values of β and $c(z)$.

We introduce the parameter $\beta(k)=\omega/k$ and write

$$Q(z,k)=\begin{cases}+k\{[\beta(k)/c(z)]^2-1\}^{\frac{1}{2}} & \beta(k)>c(z),\\ -ik\{1-[\beta(k)/c(z)]^2\}^{\frac{1}{2}} & \beta(k)<c(z).\end{cases} \tag{3.01}$$

Referring to Fig. 2, let a_1, a_2, and a_3 be the values of z at which $Q(z,k)=0$ in the ranges $0<z<z_1$, $z_1<z<z_2$, and $z_2<z$, respectively. When $\beta(k)<c_1$, only the root a_3 exists; when $c_1<\beta(k)<c_0$, all three roots exist; when $c_0<\beta(k)<c_2$, a_2 and a_3 exist; and when $\beta(k)>c_2$, Q has no zeros. It has been shown by R. E. Langer[5] that approximate solutions of Eq. (2.02), asymptotically valid for sufficiently high frequencies, may be written in the form

$$F_m=(u_m/Q)^{\frac{1}{2}}C_m(u_m), \tag{3.02}$$

where

$$u_m(z,k)=\int_{a_m}^{z}Q\,dz, \tag{3.03}$$

and C_m is a linear combination of Bessel functions of order 1/3. The solution with $m=1$ is valid in the range $0<z<a_2$ except in the immediate neighborhood of the point $z=a_1$; with $m=2$ the solution is valid for $a_1<z<a_3$ except near $z=a_2$; and with $m=3$ it is valid for $z>a_3$. The coefficients of the Bessel functions in C_m are determined by the condition that both F_1 and F_2 are asymptotic representations of the same function on the interval $a_1<z<a_2$ and similarly for F_2 and F_3 on the interval $a_2<z<a_3$. If either $c(z)$ or $c'(z)$ had a discontinuity, it would be necessary to allow the coefficients of the Bessel functions to change discontinuously at such points, corresponding physically to the existence of a reflected wave. We do not wish to complicate the present discussion by a consideration of such effects and it is for that reason that the continuity of c and c' was assumed above.

For the purpose of writing explicit expressions for the function $N(z,k)$ we divide the (β, z) plane into domains as shown in Fig. 2. Recalling that for very large altitudes N is to represent upward traveling waves, the appropriate expression for this function in domains A, B, and C is

$$N=(u_3/Q)H_{\frac{1}{3}}^{(2)}(u_3). \tag{3.04}$$

[5] R. E. Langer, Phys. Rev. **51**, 669 (1937).

Since $H_{\frac{1}{3}}^{(2)}$ is a multiple-valued function of u, it is necessary to consider the way in which the phase angles to be assigned to the u's vary between the different domains. As indicated by Eq. (3.01), $\arg Q=0$ in domains A, D, and G, while in domains B, C, and E, $\arg Q=-\pi/2$. In the neighborhood of a_3, Q is proportional to $(z-a_3)^{\frac{1}{2}}$, so that u_3 is proportional to $(z-a_3)^{\frac{1}{2}}$ or Q^3. Thus, $\arg u_3=3\arg Q$. Near a_2, Q is proportional to $(a_2-z)^{\frac{1}{2}}$ and u_2 is proportional to $-Q^3$, so that $\arg u_2=3\arg Q+\pi$. Near a_1 we again have $\arg u_1=3\arg Q$. The values to be assigned to $\arg u$ in the various domains are then given by Table I. In domain G, Q has no zeros and we may choose a arbitrarily. In this case N is not represented by the same combination of Bessel functions on opposite sides of a, but the leading term in its asymptotic expansion is the same on both sides. For continuity with domain A we set $a=z_2$ in domain G.

The leading terms in the asymptotic expansions of $H_{\frac{1}{3}}^{(1)}(u)$ and $H_{\frac{1}{3}}^{(2)}(u)$ for the values of $\arg u$ given in Table I are as follows:[6]

$\arg u=-\pi/2$, 0, or π

$$H_{\frac{1}{3}}^{(1)}(u)\rightarrow(2/\pi u)^{\frac{1}{2}}\exp[i(u-5\pi/12)], \tag{3.05}$$

$\arg u=-3\pi/2$

$$H_{\frac{1}{3}}^{(1)}(u)\rightarrow(2/\pi u)^{\frac{1}{2}}\{\exp[i(u-5\pi/12)]+\exp[-i(u+11\pi/12)]\}, \tag{3.06}$$

$\arg u=-3\pi/2$, $-\pi/2$, or 0

$$H_{\frac{1}{3}}^{(2)}(u)\rightarrow(2/\pi u)^{\frac{1}{2}}\exp[-i(u-5\pi/12)], \tag{3.07}$$

$\arg u=\pi$

$$H_{\frac{1}{3}}^{(2)}(u)\rightarrow(2/\pi u)^{\frac{1}{2}}\{\exp[i(u+11\pi/12)]+\exp[-i(u-5\pi/12)]\}. \tag{3.08}$$

We also define the following positive real quantities:

$$\left.\begin{aligned} y_1(z)&=-iu_1/k=\int_z^{a_1}\{1-[\beta(k)/c(z)]^2\}^{\frac{1}{2}}dz && 0<z<a_1\\ y_2(z)&=iu_2/k=\int_{a_2}^{z}\{1-[\beta(k)/c(z)]^2\}^{\frac{1}{2}}dz && a_2<z<a_3\\ y_3(z)&=-iu_3/k=\int_z^{a_3}\{1-[\beta(k)/c(z)]^2\}^{\frac{1}{2}}dz && a_2<z<a_3\end{aligned}\right\}. \tag{3.09}$$

$$\left.\begin{aligned} x_1(z)&=u_1/k=\int_{a_1}^{z}\{[\beta(k)/c(z)]^2-1\}^{\frac{1}{2}}dz && a_1<z<a_2\\ x_2(z)&=-u_2/k=\int_z^{a_2}\{[\beta(k)/c(z)]^2-1\}^{\frac{1}{2}}dz && a_1<z<a_2\\ x_3(z)&=u_3/k=\int_{a_3}^{z}\{[\beta(k)/c(z)]^2-1\}^{\frac{1}{2}}dz && a_3<z\end{aligned}\right\}. \tag{3.10}$$

[6] G. N. Watson, *Bessel Functions*, second edition (Cambridge University Press, London, 1944), pp. 75–201.

TABLE I.

Domain	arg$Q$	argu_1	argu_2	argu_3
A	0	...	...	0
B	$-\pi/2$	...	...	$-3\pi/2$
C	$-\pi/2$	...	$-\pi/2$	$-3\pi/2$
D	0	0	π	...
E	$-\pi/2$	$-3\pi/2$	...	...

From the expressions (3.05) to (3.08) and the asymptotic continuity conditions, we find the following expressions for the function $N(z, k)$ in the various domains:

In A

$$N=(u_3/Q)^{\frac{1}{2}}H_{\frac{1}{2}}^{(2)}(u_3)\rightarrow(2/\pi Q)^{\frac{1}{2}} \times\exp\{-i[kx_3(z)-5\pi/12]\}. \quad (3.11)$$

In B

$$N=(u_3/Q)^{\frac{1}{2}}H_{\frac{1}{2}}^{(2)}(u_3)\rightarrow(2/\pi Q)^{\frac{1}{2}} \times\exp[ky_3(z)+5\pi i/12]. \quad (3.12)$$

In C

$$N=(u_2/Q)^{\frac{1}{2}}\exp[ky_2(a_3)]H_{\frac{1}{2}}^{(2)}(u_2)\rightarrow(2/\pi Q)^{\frac{1}{2}} \times\exp[-ky_2(z)+ky_2(a_3)+5\pi i/12]. \quad (3.13)$$

In D

$$N=(u_2/Q)^{\frac{1}{2}}\exp[ky_2(a_3)]H_{\frac{1}{2}}^{(2)}(u_2)\rightarrow 2(2/\pi Q)^{\frac{1}{2}} \times\exp[ky_2(a_3)+2\pi i/3]\cos[kx_2(z)-\pi/4]. \quad (3.14)$$

In E

$$\begin{aligned}N=(u_1/Q)^{\frac{1}{2}}&\exp[ky_2(a_3)]\\ &\times\{\exp[-ikx_2(a_1)+4\pi i/3]H_{\frac{1}{2}}^{(1)}(u_1)\\ &+\exp[ikx_2(a_1)]H_{\frac{1}{2}}^{(2)}(u_1)\}\rightarrow(2/\pi Q)^{\frac{1}{2}}\\ &\times\exp[ky_2(a_3)+2\pi i/3]\\ &\times\{\exp[-ky_1(z)-ikx_2(a_1)+\pi i/4]\\ &+2\cos kx_2(a_1)\exp[ky_1(z)-\pi i/4]\}. \quad (3.15)\end{aligned}$$

In G

$$N\rightarrow(2/\pi Q)^{\frac{1}{2}}\exp[-i(u-5\pi/12)], \quad (3.16)$$

$$u=\int_{z_3}^{z}Q\,dz.$$

With these expressions for N, there are discontinuities in $F(z, k)$ with respect to variations of k at $k=\omega/c_1$ and ω/c_2. The transformation of the path of integration in the integral expression for φ leads, in addition to the contributions from the residues given by Eq. (2.10), to integrals to and from $\omega/c_2-i\infty$, $\omega/c_1-i\infty$, and $\omega/c(z)-i\infty$ around the cuts in the k plane shown in Fig. 3. There is no discontinuity with respect to k at the cut starting at $\omega/c(z)$ on the real axis, so that the integral around this cut vanishes. The integral around the cut starting at ω/c_1 does not vanish because of the discontinuity in F between domains B and E, but it may be shown that this integral is of order r^{-1} as compared with the residue terms which are of order $r^{-\frac{1}{2}}$. This integral is therefore negligible when r is large. Similarly, the integral around the cut starting at ω/c_2 does not vanish. It will appear presently that this integral is to be interpreted as representing the effect of rays that leave the source at angles steep enough so that they are not refracted back below the level z_2. This integral is therefore negligible for large values of r when $z<z_2$, which is the only case with which we shall be concerned; but it would not be negligible if we wished to compute values of φ for $z>z_2$.

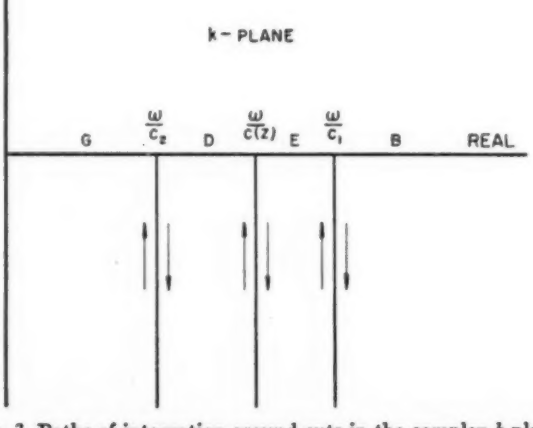


FIG. 3. Paths of integration around cuts in the complex k plane.

4. THE ROOTS OF $N'(0, k)$

We have

$$\begin{aligned}d[(u/Q)^{\frac{1}{2}}H_{\frac{1}{2}}(u)]/dz&\\ =[(1/6)(Q/u)^{\frac{1}{2}}-(1/2)(u/Q)^{\frac{1}{2}}(Q'/Q)]H_{\frac{1}{2}}(u)&\\ +(uQ)^{\frac{1}{2}}H_{-\frac{1}{2}}(u), \quad (4.01)&\end{aligned}$$

where H stands for either of the Bessel functions $H^{(1)}$ or $H^{(2)}$. For a given value of $\beta(k)$ the expression in brackets in Eq. (4.01) is independent of ω, while $(uQ)^{\frac{1}{2}}$ is proportional to ω; hence, we may write

$$d[(u/Q)^{\frac{1}{2}}H_{\frac{1}{2}}]/dz\simeq(uQ)^{\frac{1}{2}}H_{-\frac{1}{2}}(u), \quad (4.02)$$

to the same approximation as that involved in writing N in terms of $H_{\frac{1}{2}}$. Since the leading term in the asymptotic expansion of $H_{-\frac{1}{2}}$ may be obtained from that for $H_{\frac{1}{2}}$ by multiplying the $\exp(iu)$ term by $\exp(\pi i/2)$ and the $\exp(-iu)$ term by $\exp(-\pi i/2)$ for any value of $\arg u$, the asymptotic expressions for $N(0, k)$ in the various domains are readily derived from those for N by inspection.

In B

$$N'(0,k)=[u_3(0)Q(0)]^{\frac{1}{2}}H_{-\frac{1}{2}}^{(2)}[u_3(0)]\rightarrow[2Q(0)/\pi]^{\frac{1}{2}} \times\exp[ky_3(0)-\pi i/12]. \quad (4.03)$$

In E

$$\begin{aligned}N'(0,k)=[u_1(0)Q(0)]^{\frac{1}{2}}&\exp[ky_2(a_3)]\\ &\times\{\exp[-ikx_2(a_1)+4\pi i/3]H_{-\frac{1}{2}}^{(1)}[u_1(0)]\\ &+\exp[ikx_2(a_1)]H_{-\frac{1}{2}}^{(2)}[u_1(0)]\}\rightarrow[2Q(0)/\pi]^{\frac{1}{2}}\\ &\times\exp[ky_2(a_3)+2\pi i/3]\\ &\times\{\exp[-ky_1(0)-ikx_2(a_1)+3\pi i/4]\\ &+2\cos kx_2(a_1)\exp[ky_1(0)-3\pi i/4]\}. \quad (4.04)\end{aligned}$$

In D

$$\begin{aligned}N'(0,k)=[u_2(0)Q(0)]^{\frac{1}{2}}&H_{-\frac{1}{2}}^{(2)}[u_2(0)]\\ &\rightarrow 2[2Q(0)/\pi]^{\frac{1}{2}}\exp[ky_2(a_3)+2\pi i/3]\\ &\times\cos[kx_2(0)-3\pi/4]. \quad (4.05)\end{aligned}$$

In G

$$N'(0,k)\rightarrow[2Q(0)/\pi]^{\frac{1}{2}}\exp\{-i[u(0)+\pi/12]\}. \quad (4.06)$$

From Eq. (4.03) we see that $N'(0,k)$ has no zeros for $\beta(k)<c_1$, so that the series (2.10) begins with values of k_n such that $\beta(k_n)>c_1$. From Eq. (4.04) it appears that there are no real roots in the range $c_1<|\beta(k_n)|<c_0$. However, it is to be noted that the asymptotic expansions of the Bessel functions are valid only when the quantities kx and ky are large. Consequently, $\exp[-ky_1(0)]$ is very small compared with $\exp[ky_1(0)]$, and the values of k_n in this range will be very nearly given by $\cos k_n x_2(a_1)=0$ or

$$k_n x_2(a_1)\simeq(n-\tfrac{1}{2})\pi, \quad (4.07)$$

where n is a positive integer. From Eq. (4.05) the roots in the range $c_0<\beta(k)<c_2$ are given by

$$k_n x_2(0)=(n-\tfrac{3}{4})\pi, \quad (4.08)$$

and from Eq. (4.06) there are no roots for $\beta(k)>c_2$.

Equations (4.07) and (4.08) for k_n, being derived from the asymptotic expansions are, of course, not valid for small values of n, although they are correct to three significant figures or better for $n>4$. Explicit expressions for the lower order roots in domain E cannot be readily derived; but in domain D, they may be computed from the equation

$$H_{-\frac{1}{3}}^{(2)}[u_2(0)]=(2/\sqrt{3})\times\exp(\pi i/6)\{J_{\frac{1}{3}}[kx_2(0)]-J_{-\frac{1}{3}}[kx_2(0)]\}. \quad (4.09)$$

The first five roots of this expression are $k_n x_2(0)=0.6855$, 3.9028, 7.0549, 10.2007, and 13.3445. In the case of a typical atmospheric velocity-altitude structure the sum (2.10) for φ includes a large number of high order modes when the frequency is greater than about 0.2 cycle, and the contribution from the few low order modes is comparatively small. For very much lower frequencies the only normal mode solutions that exist are those of low order. However, at such frequencies the representation of the function $N(z,k)$ in terms of Bessel functions becomes a poor approximation. The present theory may therefore be considered valid only for frequencies sufficiently high that the major contribution to φ comes from modes of order $n>4$. For these modes, the asymptotic forms of the Bessel functions provide a sufficiently accurate representation, and these forms will accordingly be used in all subsequent development.

5. ASYMPTOTIC APPROXIMATION FOR ϕ

We now have explicit expressions for all the quantities appearing in Eq. (2.10) except the factor $[\partial N'(0,k)/\partial k]_{k=k_n}$.

From the definition (3.14)

$$\partial[kx_2(a_1)]/\partial k=-k\int_{a_1}^{a_2}Q^{-1}dz \quad c_1<\beta(k)<c_0. \quad (5.01)$$

For $\beta(k)>c_0$, a_1 is replaced by 0 in this expression.

From Eqs. (4.04), (4.05), and (5.01) with the condition $N'(0,k_n)=0$, we find

$$[\partial N'(0,k)/\partial k]_{k=k_n}\rightarrow 2k_n(-1)^{n-1}\times\exp[k_n y_2(a_3)+k_n y_1(0)-\pi i/12]\times(2Q_n(0)/\pi)^{\frac{1}{2}}\int_{a_1}^{a_2}Q_n^{-1}dz, \quad (5.02)$$

for $c_1<\beta(k_n)<c_0$, and

$$[\partial N'(0,k)/\partial k]_{k=k_n}\rightarrow 2k_n(-1)^{n}\times\exp[k_n y_2(a_3)+2\pi i/3]\times(2Q_n(0)/\pi)^{\frac{1}{2}}\int_{0}^{a_2}Q_n^{-1}dz, \quad (5.03)$$

for $c_0<\beta(k_n)<c_2$. From Eq. (3.20) with Eq. (5.02) and from Eq. (3.10) with Eq. (5.03), we have

$$N(0,k_n)[\partial N'(0,k)/\partial k]_{k=k_n}\rightarrow(8k_n/\pi)\times\exp[2k_n y_2(a_3)+\pi i/3]\int_{a_1}^{a_2}Q_n^{-1}dz, \quad (5.04)$$

for $c_1<\beta(k_n)<c_0$ and the same expression with 0 replacing a_1 for $c_0<\beta(k_n)<c_2$.

We shall assume the source to be at an altitude less than that of the velocity minimum, that is $h<z_1$, so that $N(h,k_n)$ is given by Eqs. (3.18) and (3.19)

$$N(h,k_n)\rightarrow\begin{cases}2[2/\pi Q_n(h)]^{\frac{1}{2}}(-1)^{n+1}\times\exp[k_n y_2(a_3)-k_n y_1(0)+5\pi i/12]\times\cosh\{k_n[y_1(h)-y_1(0)]\} & c_1<\beta(k_n)<c(h), \quad (5.05)\\ 2[2/\pi Q_n(h)]^{\frac{1}{2}}\exp[k_n y_2(a_3)+2\pi i/3]\times\cos[k_n x_2(h)-\pi/4] & c(h)<\beta(k_n)<c_2.\end{cases}$$

Substituting Eqs. (5.04) and (5.05) in Eq. (2.10) and using the asymptotic form of $H_0^{(2)}(k_n r)$, we may write

$$\phi_{res}=\phi_1+\phi_2+\phi_3,$$

where

$$\phi_1=\pi A\exp[i(\omega t-\pi/6)]\sum_{c_1<\beta(k_n)<c(h)}(-1)^{n+1}\times[rk_nQ_n(h)]^{-\frac{1}{2}}\exp[-ik_n r-k_n y_2(a_3)-k_n y_1(0)]\times\cosh\{k_n[y_1(h)-y_1(0)]\}\times N(z,k_n)\Big/\int_{a_1}^{a_2}Q_n^{-1}dz, \quad (5.06)$$

$$\phi_2=\pi A\exp[i(\omega t+\pi/12)]\sum_{c(h)<\beta(k_n)<c_0}[rk_nQ_n(h)]^{-\frac{1}{2}}\times\exp[-ik_n r-k_n y_2(a_3)]\cos[k_n x_2(h)-\pi/4]\times N(z,k_n)\Big/\int_{a_1}^{a_2}Q_n^{-1}dz, \quad (5.07)$$

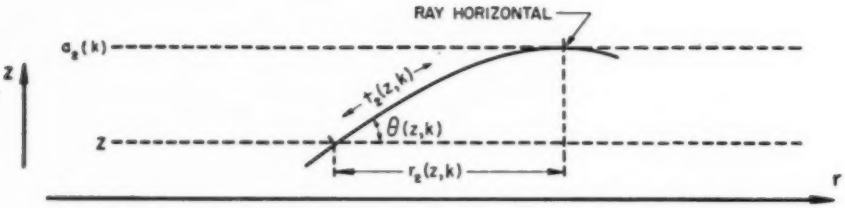


FIG. 4. Ray paths in the (r, z) plane.

$$\phi_3 = \pi A \exp[i(\omega t+\pi/12)] \sum_{c_0<\beta(k_n)<c_2} [rk_nQ_n(h)]^{-\frac{1}{2}} \times\exp[-ik_nr-k_ny_2(a_3)]\cos[k_nx_2(h)-\pi/4] \times N(z, k_n)\Big/\int_0^{a_2} Q_n^{-1}dz. \quad (5.08)$$

Now each term of φ_1 contains the factors $\exp[-k_ny_2(a_3)]$ and $\exp[-k_ny_1(0)]$, both of which are very small. Except when $[z, \beta(k_n)]$ falls in domain A, the factor $\exp[-ky_2(a_3)]$ is cancelled by its inverse in $N(z, k_n)$; but the factor $\exp[-k_ny_1(0)]$ remains in all cases. Thus, ϕ_1 is negligibly small for all values of z. φ_2 is likewise negligible owing to the factor $\exp[-k_ny_2(a_3)]$ when $[z, \beta(k_n)]$ is in domain A, and all terms in φ_2 for which $[z, \beta(k_n)]$ falls in domain C are also negligible owing to the factor $\exp[-k_ny_2(z)]$, which appears in $N(z, k_n)$ in this domain. When $[z, \beta(k_n)]$ falls in domain E, Eq. (3.20) gives

$$N(z, k)\rightarrow 2[2/\pi Q_n(z)]^{\frac{1}{2}}(-1)^{n+1} \times\exp[k_ny_2(a_3)-k_ny_1(0)+5\pi i/12] \times\cosh\{k_n[y_1(z)-y_1(0)]\},$$

and the corresponding terms of ϕ_2 are negligible owing to the factor $\exp[-k_ny_1(0)]$. The surviving terms of φ_2 are, therefore, those for which $[z, \beta(k_n)]$ lies in domain D. For these we have from Eq. (3.19)

$$\phi_2 = 2A \exp[i(\omega t+3\pi/4)] \sum_{c(h)<\beta(k_n)<c_0} [2\pi/rk_nQ_n(h)Q_n(z)]^{\frac{1}{2}} \times\exp(-ik_nr)\cos[k_nx_2(h)-\pi/4] \times\cos[k_nx_2(z)-\pi/4] \times\delta[z, a_1(k_n), a_2(k_n)]\Big/\int_{a_1}^{a_2} Q_n^{-1}dz, \quad (5.09)$$

where $\delta[z, a_1(k_n), a_2(k_n)]$ is a factor that has the value 1 for $a_1(k_n)<z<a_2(k_n)$ and 0 for z outside this range. A similar argument leads to

$$\phi_3 = 2A \exp[i(\omega t+3\pi/4)] \sum_{c_0<\beta(k_n)<c_2} [2\pi/rk_nQ_n(h)Q_n(z)]^{\frac{1}{2}} \times\exp(-ik_nr)\cos[k_nx_2(h)-\pi/4] \times\cos[k_nx_2(z)-\pi/4] \times\delta[z, 0, a_2(k_n)]\Big/\int_0^{a_2} Q^{-1}dz. \quad (5.10)$$

In Eqs. (5.09) and (5.10) the values of k_n are to be determined by Eqs. (4.07) and (4.08), respectively.

6. GEOMETRICAL INTERPRETATION OF THE NORMAL MODES

It will be shown presently that the summations indicated in Eqs. (5.09) and (5.10) may be carried out by replacing the sums by equivalent integrals, which may then be evaluated approximately by the method of stationary phase. However, before proceeding with this development, it will be instructive to consider the significance in terms of ray geometry of the functions $Q(z, k)$ and $x_2(z, k)$ that express the dependence of ϕ on z. Let $\theta(z)$ be the angle between a given geometrical ray and the horizontal. Then Snell's law is $c(z)\sec\theta(z)$ = constant. Since the function $x_2(z)$ is defined only for values of $\beta(k)>c(z)$, $\beta(k)$ is a possible ray constant and defines a ray for which $\sec\theta(z, k)=\beta(k)/c(z)$. Equation (3.01) for this case may be written in the form

$$Q(z, k)=k[\sec^2\theta(z, k)-1]^{\frac{1}{2}}=k\tan\theta(z, k). \quad (6.01)$$

The zeros of $Q(z, k)$ are thus the altitudes at which the corresponding ray is horizontal. From the second of

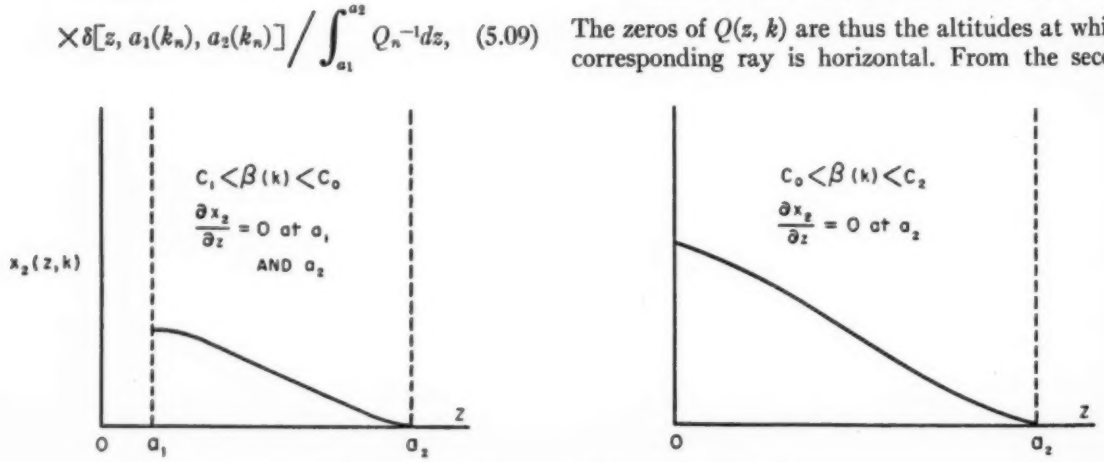


FIG. 5. General character of the function $x_2(z, k)$.

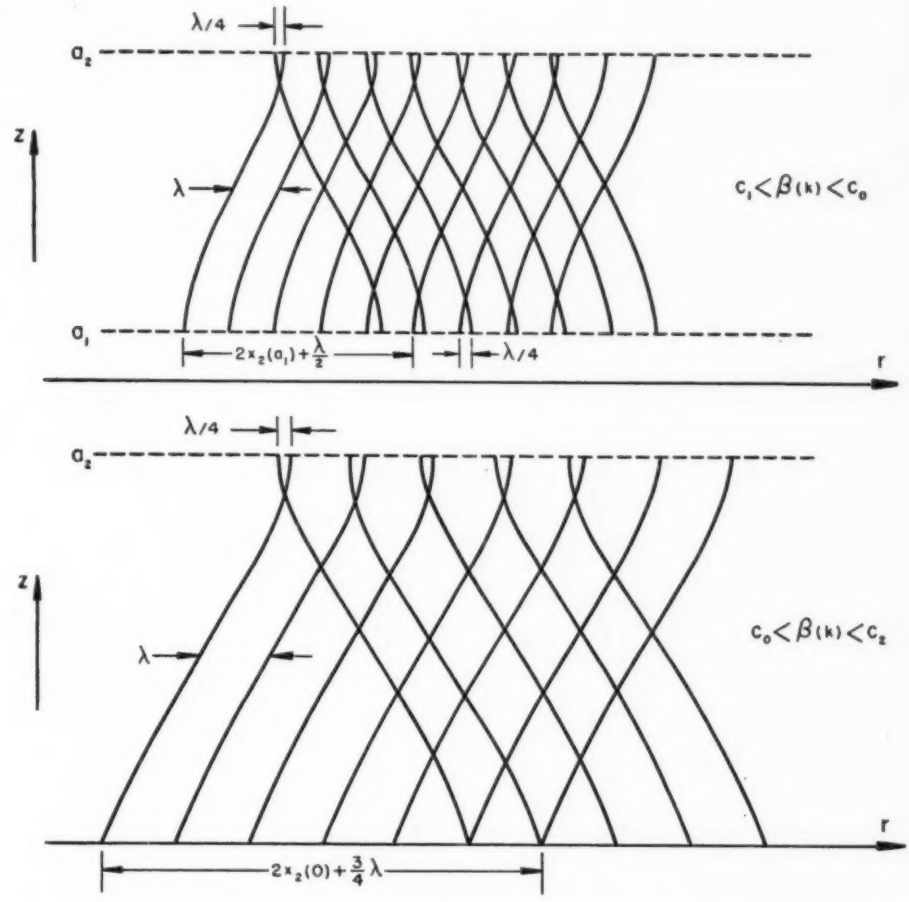


FIG. 6. Configuration of normal mode wave fronts.

Eqs. (3.14) we have

$$x_2(z)=\int_z^{a_2}\tan\theta(z,k)dz,$$

and when we write $\tan\theta=[\beta(k)/c(z)]\csc\theta-\cot\theta$, this takes the form

$$x_2(z)=\beta(k)\int_z^{a_2}[\csc\theta(z,k)/c(z)]dz-\int_z^{a_2}\cot\theta(z,k)dz.$$

The first integral in this expression is the travel time along the given ray from the altitude z to the apex of the ray at altitude a_2, and the second is the horizontal distance traversed (see Fig. 4). Calling these quantities $t_2(z,k)$ and $r_2(z,k)$, respectively, we write

$$x_2(z,k)=\beta(k)t_2(z,k)-r_2(z,k). \tag{6.02}$$

Furthermore,

$$\int_{a_1}^{a_2}Q^{-1}dz=k^{-1}\int_{a_1}^{a_2}\cot\theta dz=r_2(a_1,k)/k. \tag{6.03}$$

The general character of the function $x_2(z,k)$ for the type of velocity altitude function assumed is shown in Fig. 5.

By writing the factor $\cos[kx_2(z)-\pi/4]$ in exponential form, it may be seen that each term in Eqs. (5.09) and (5.10) represents a superposition of a pair of progressive waves whose phases are given (apart from a constant term) by

$$\omega\{t-[r-x_2(z,k_n)]/\beta(k_n)\}-\pi/4,$$

and

$$\omega\{t-[r+x_2(z,k_n)]/\beta(k_n)\}+\pi/4.$$

Thus, the parameter $\beta(k_n)$ is to be interpreted as the phase velocity of the nth normal mode, and $k_n=2\pi/\lambda_n$, where λ_n is the corresponding wavelength (measured horizontally). The configuration of the wave fronts, or loci of constant phase, is illustrated in Fig. 6. This configuration is such that the velocity of propagation of any portion of a wave front measured along the normal is $c(z)$, while the velocity measured horizontally is the same for all values of z and is equal to $\beta(k_n)$.

The conditions (4.07) and (4.08) that determine the k_n's may be written in the form

$$2x_2(a_1)=\lambda_n(n-\tfrac{1}{2}) \quad c(h)<\beta(k_n)<c_0, \tag{6.04}$$

$$2x_2(0)=\lambda_n(n-\tfrac{3}{4}) \quad c_0<\beta(k_n)<c_2. \tag{6.05}$$

It is seen from Fig. 6 that these are the conditions that must be satisfied if each of the two sets of wave fronts is to be in phase with the reflected or refracted image of the other, with allowance for a quarter wavelength phase shift when the wave front is reversed by refraction

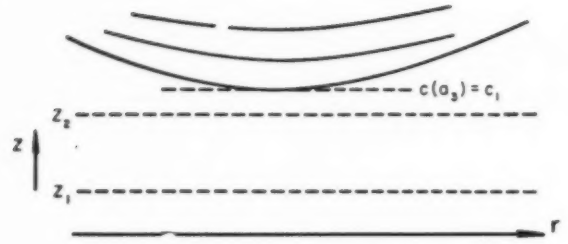


FIG. 7. Rays for which $0<\beta(k)\leqslant c_1$.

and no phase shift when the reversal is by reflection at the ground surface. This is exactly analogous to the condition for constructive interference between multiple internal reflections that determine the normal modes in the case of a medium made up of discrete layers.[1]

We also observe that the (asymptotic) vanishing of φ_1, which corresponds to the nonexcitation of the normal modes for which $\beta(k_n)<c(h)$, is the expression in the normal mode representation of the fact that the corresponding rays do not pass through the level of the source. The interpretation of $\beta(k)$, for real values of k, as a quantity defining a family of rays in the (r, z) plane indicates the physical meaning of the integrals around the cuts in the k plane. The integral around the cut starting at ω/c_1 arose by transformation of the original path of integration along the real axis for $0<\beta(k)<c_1$. The corresponding rays are confined to the portion of the (r, z) plane for which z is greater than the altitude a_3 (see Fig. 7). Such rays do not, therefore, contribute appreciably to the value of φ when either h or z is less than a_3, which is the case in which we are primarily interested. Similarly, the integral around the cut starting at ω/c_2 represents the contribution to φ of rays for which $\beta(k)>c_2$, that is, of the type shown in Fig. 8. These rays, after once passing above the level z_2, do not thereafter return to lower elevations and thus do not contribute appreciably to φ when the point (r, z) lies to the right of the upward refracted branch of the critical ray that becomes horizontal at the altitude z_2. For such points φ_{res} therefore represents essentially the whole solution.

7. PHASE AND GROUP VELOCITIES OF THE NORMAL MODES

Since, for a given value of n, $\beta(k_n)$ is a function of frequency defined by Eqs. (4.07) or (4.08), the individual terms of Eqs. (5.09) and (5.10) represent dispersed waves. It is, therefore, of interest to compute the corresponding group velocities. The group velocity U_n corresponding to the phase velocity $\beta_n=\beta(k_n)$ is defined by

$$U_n=d\omega/dk_n. \tag{7.01}$$

From Eq. (4.07) we have for $n=$ constant

$$x_2(a_1)dk_n+k_n dx_2(a_1)=0,$$

or

$$\{x_2(a_1)+k_n[\partial x_2(a_1)/\partial k_n]_\omega\}dk_n + k_n[\partial x_2(a_1)/\partial\omega]_{k_n}d\omega=0. \tag{7.02}$$

From

$$x_2(a_1)=k_n^{-1}\int_{a_1}^{a_2}\{[\omega/c(z)]^2-k_n^2\}^{-\frac{1}{2}}dz,$$

and the fact that the integrand vanishes at a_1 and a_2 we have

$$[\partial x_2(a_1)/\partial\omega]_{k_n}=k_n^{-1}\int_{a_1}^{a_2}[c(z)]^{-1} \times\{1-[c(z)/\beta(k_n)]^2\}^{-\frac{1}{2}}dz=t_2(a_1)/k_n, \tag{7.03}$$

and

$$[\partial x_2(a_1)/\partial k_n]_\omega=-k_n^{-1}\int_{a_1}^{a_2}\tan\theta dz-k_n^{-1}\int_{a_1}^{a_2}\cot\theta dz =-k_n^{-1}[x_2(a_1)+r_2(a_1)] =-\beta_n t_2(a_1)/k_n, \tag{7.04}$$

whence

$$U_n=r_2(a_1)/t_2(a_1). \tag{7.05}$$

Equation (4.08) leads to the same relations with a_1 replaced by 0. Thus, the group velocity of the nth mode is equal to the average horizontal velocity along the corresponding ray between its maximum and minimum altitudes.

A numerical example is plotted in Figs. 9(a) and 9(b), which show β_n and U_n as functions of frequency for the following velocity-altitude function:

Alt. (km)	c (meters/sec)
0	344
16.5	286
32	312
50	380

Linear velocity gradients are assumed between the tabulated values and the velocity is assumed to decrease monotonically above 50 kilometers. The fact that this velocity function does not fulfill the condition of continuity of dc/dz is immaterial for our present purpose, since the ray geometry would not be greatly altered if we supposed the discontinuities in gradient to be rounded off to satisfy the continuity condition. The phase velocity curves provide the basis for the statement previously made that for frequencies higher than about 0.2 cycle φ includes a large number of high order modes. For example, the number of terms in φ_3 for any given frequency is equal to the number of values of β_n at that frequency falling between the limits c_0 and c_3.

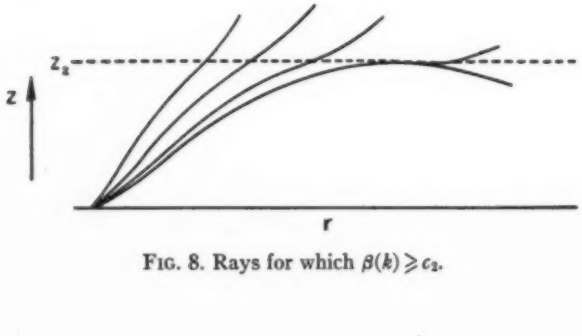


FIG. 8. Rays for which $\beta(k)\geqslant c_2$.

Thus, at 0.2 cycle there are twelve terms, from the 22nd to the 33rd mode inclusive, at 1.0 cycle there are fifty-eight terms, from the 107th to the 164th mode, and at 5.0 cycles there are two hundred and eighty-eight terms, from the 532nd to the 819th mode.

8. INTEGRAL REPRESENTATION OF ϕ

Since the number of terms is so large, Eqs. (5.09) and (5.10) are not very useful as they stand. However, the fact that successive values of β_n differ only slightly suggests that the sums may be converted into equivalent integrals. Using the symbol Δ to denote the change in any quantity from the nth to the $(n+1)$th mode, we have from either of Eqs. (4.07) or (4.08)

$$\Delta[k_n x_2(a_1)]/\pi=\Delta n=1.$$

But

$$\Delta[k_n x_2(a_1)]\simeq\Delta k_n\{x_2(a_1)+k_n[\partial x_2(a_1)/\partial k_n]_\omega\}$$
$$=-r_2(a_1)\Delta k_n,$$

from Eq. (7.04). Hence,

$$-r_2(a_1)\Delta k_n/\pi=1. \tag{8.01}$$

Multiplying each term of Eqs. (5.09) and (5.10) by Eq. (8.01) and using Eqs. (6.01) and (6.03), we have

$$\phi_2=2A\exp[i(\omega t-\pi/4)]\sum_{c(h)<\beta_n<c_0}[2/\pi r k_n\tan\theta_n(z)\tan\theta_n(h)]^{\frac{1}{2}}$$
$$\times\exp[-ik_n r]\cos[k_n x_2(h)-\pi/4]$$
$$\times\cos[k_n x_2(z)-\pi/4]\delta(z, a_1, a_2)\Delta k_n, \tag{8.02}$$

$$\phi_3=2A\exp[i(\omega t-\pi/4)]\sum_{c_0<\beta_n<c_2}[2/\pi r k_n\tan\theta_n(z)\tan\theta_n(h)]^{\frac{1}{2}}$$
$$\times\exp[-ik_n r]\cos[k_n x_2(h)-\pi/4]$$
$$\times\cos[k_n x_2(z)-\pi/4]\delta(z, 0, a_2)\Delta k_n. \tag{8.03}$$

Equation (8.01) shows that when $r_2(a_1)$ is a large number of wavelengths, Δk_n may be treated as a small quantity. However, before passing to the limit $\Delta k_n\to dk_n$ and writing these sums in the form of integrals over a continuous range of values of k_n, it is necessary to take account of the fact that in deriving Eqs. (5.09) and (5.10) terms depending upon $x_2(a_1)$ and $x_2(0)$ were eliminated by use of the relations (4.07) and (4.08). Before allowing k_n to assume continuous values we reintroduce the dependence upon $x_2(a_1)$ and $x_2(0)$ by noting that for any integer, s,

$$\exp\{-2is[k_n x_2(a_1)+\pi/2]\}=\exp[-2i\pi ns]=1,\quad c(h)<\beta_n<c_0,$$
$$\exp\{-2is[k_n x_2(0)+3\pi/4]\}=\exp[-2i\pi ns]=1,\quad c_0<\beta_n<c_2.$$

These expressions may, therefore, be introduced as factors after the summation signs in Eqs. (8.02) and (8.03). The reason for the introduction of the arbitrary integer s will become apparent presently. Passing now to the limit $\Delta k_n\to dk$ and writing the cosine terms in exponential form, Eqs. (8.02) and (8.03) become

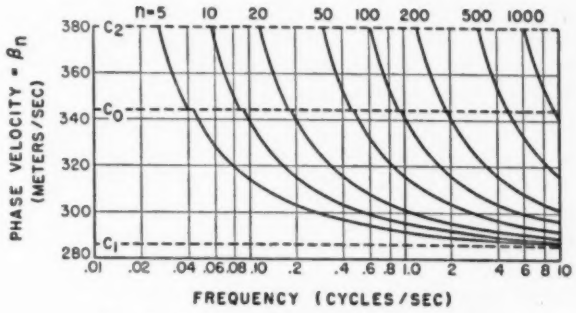


FIG. 9a. Phase velocity of normal modes.

$$\phi_2=(A/2)\exp[i(\omega t-\pi/4-s\pi)]$$
$$\times[I_1\exp(-\pi i/2)+I_2\exp(\pi i/2)+I_3+I_4], \tag{8.04}$$

$$\phi_3=(A/2)\exp[i(\omega t-\pi/4-3s\pi/2)]$$
$$\times[I_5\exp(-\pi i/2)+I_6\exp(\pi i/2)+I_7+I_8], \tag{8.05}$$

where

$$I_1=\int_{k_h}^{k_0^+}f(k)\delta(z, a_1, a_2)$$
$$\times\exp\{-ik[r-x_2(h)-x_2(z)+2sx_2(a_1)]\}dk, \tag{8.06}$$

$$I_2=\int_{k_h}^{k_0^+}f(k)\delta(z, a_1, a_2)$$
$$\times\exp\{-ik[r+x_2(h)+x_2(z)+2sx_2(a_1)]\}dk, \tag{8.07}$$

$$I_3=\int_{k_h}^{k_0^+}f(k)\delta(z, a_1, a_2)$$
$$\times\exp\{-ik[r-x_2(h)+x_2(z)+2sx_2(a_1)]\}dk, \tag{8.08}$$

$$I_4=\int_{k_h}^{k_0^+}f(k)\delta(z, a_1, a_2)$$
$$\times\exp\{-ik[r+x_2(h)-x_2(z)+2sx_2(a_1)]\}dk, \tag{8.09}$$

$$f(k)=[2/\pi rk\tan\theta(z, k)\tan\theta(h, k)]^{\frac{1}{2}}. \tag{8.10}$$

The lower limit of integration, k_h, is the first value of k_n less than $\omega/c(h)$ and k_0^+ is the value of k_n next larger than ω/c_0. The integrals I_5 to I_8 are given by the same

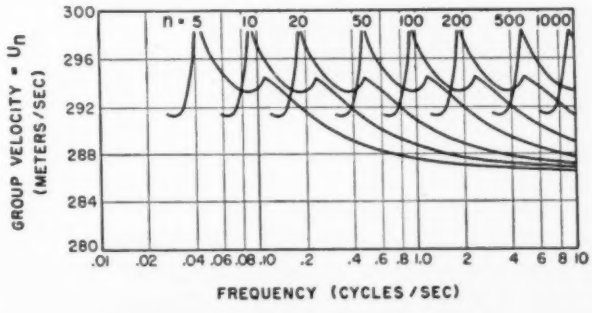


FIG. 9b. Group velocity of normal modes.

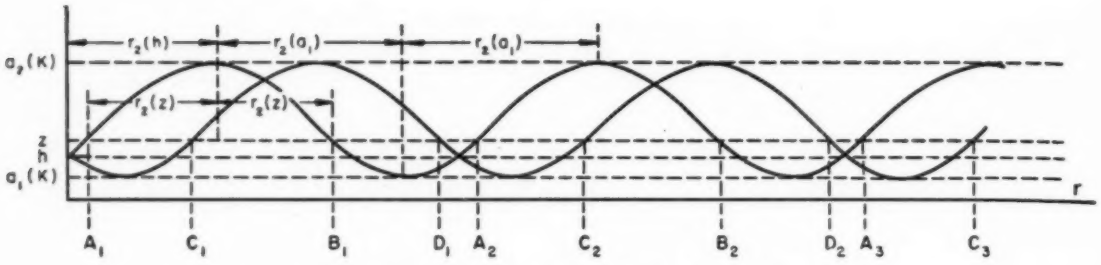


FIG. 10. Ray geometry and points of stationary phase.

expressions with $a_1=0$ and the lower and upper limits of integration replaced by k_0^- and k_2, respectively, where k_0^- is the value of k_n next smaller than ω/c_0 and k_2 is the value next larger than ω/c_2.

9. APPROXIMATE EVALUATION OF THE INTEGRALS

In the approximate evaluation of integrals of the type (8.06)–(8.09), it is necessary to consider the criteria for the existence of stationary points and points of inflection of the phase of the exponential factors in the integrands. We let $X(z, h, k)$ stand for any one of the four quantities $2sx_2(a_1)\pm x_2(z)\pm x_2(h)$, so that the phase may be written in the form

$$\psi(r, z, h, k)=k[r+X(z, h, k)]. \tag{9.01}$$

We also introduce the notation $R(z, h, k)=2sr_2(a_1)\pm r_2(z)\pm r_2(h)$ and $T(z, h, k)=2st_2(a_1)\pm t_2(z)\pm t_2(h)$. Then, using Eqs. (5.01) and (6.03) (which remain true if z or h is substituted for the lower limit of integration, a_1), we have

$$\partial\psi/\partial k=r-R(z, h, k), \tag{9.02}$$

$$\partial^2\psi/\partial k^2=-\partial R/\partial k, \tag{9.03}$$

$$\partial^3\psi/\partial k^3=-\partial^2 R/\partial k^2. \tag{9.04}$$

The condition for a point of stationary phase is then $r=R(z, h, k)$, which is to be regarded as an equation to determine a particular value (or set of values) of k for given values of r, z, and h. The geometrical significance of this condition is illustrated in Fig. 10, which shows a number of refracted cycles of the pair of rays specified by a particular value of k, say K. The radial distances to the points A_1, A_2, $\cdots$, B_1, B_2, etc., at altitude z are given by the equations,

$$\left.\begin{aligned} r_A&=2sr_2(a_1)+r_2(h)-r_2(z) && s=0, 1, 2, \cdots\\ r_B&=2sr_2(a_1)+r_2(h)+r_2(z) && s=0, 1, 2, \cdots\\ r_C&=2sr_2(a_1)-r_2(h)-r_2(z) && s=1, 2, \cdots\\ r_D&=2sr_2(a_1)-r_2(h)+r_2(z) && s=1, 2, \cdots \end{aligned}\right\}, \tag{9.05}$$

all of which are comprised in the general condition for the existence of a point of stationary phase for one or more of the integrals (8.06)–(8.09). Thus, we see that whenever the point of observation (r, z) lies on a possible geometrical ray from the source, at least one of the integrals I_1, $\cdots$ I_8 will have a point of stationary phase for some value of s. If (r, z) falls at a point where two rays having different values of K intersect, there will be two points of stationary phase; and if (r, z) lies in a geometrical shadow zone, the equation $r=R(z, h, k)$ will have no solution for real values of k and none of the integrals will have points of stationary phase. The integer s is now seen to represent the number of times the given ray passes through its minimum altitude (or is reflected at the ground surface) in reaching the point (r, z).

Approximate evaluations[7] of the general integral appropriate under various circumstances are as follows:

Case I. (r, z) does not lie in a shadow zone and is not a focal point, so that $r=R(K)$; but $\partial R/\partial K\neq 0$.

$$I\simeq[-\pi/(\partial R/\partial K)]^{\frac{1}{2}}f(K)\exp[-iK\{r+X(K)\}]\times\int_{v_a}^{v_b}\exp(-i\pi v^2/2)dv, \tag{9.06}$$

where

$$v_b=+[(2/\pi)\{K[r+X(K)]-k_b[r+X(k_b)]\}]^{\frac{1}{2}}, \tag{9.07}$$

$$v_a=-[(2/\pi)\{K[r+X(K)]-k_a[r+X(k_a)]\}]^{\frac{1}{2}}. \tag{9.08}$$

In the integrals I_1 to I_4, $k_b=k_0^+$, $k_a=k_h$, and in the integrals I_5 to I_8, $k_b=k_2$, $k_a=k_0^-$.

If the approximate evaluation of I represented by Eq. (9.06) is carried further, the next term is

$$i\{[f(k_b)-f(K)]/[r-R(k_b)]\}\exp\{-ik_b[r+X(k_b)]\},$$
$$-i\{[f(k_a)-f(K)]/[r-R(k_a)]\}\exp\{-ik_a[r+X(k_a)]\},$$

and this term should be small if Eq. (9.06) is to be used as it stands. For a given value of $\beta(k)$, $f(k)$, and $\partial R/\partial k$ are proportional to $\omega^{-\frac{1}{2}}$, so that, apart from the fluctuations due to the presence of the fresnel integral, the absolute value of the first term is independent of the frequency and the second is proportional to $\omega^{-\frac{1}{2}}$. However, in addition to this explicit dependence on the frequency, there is an implicit dependence owing to the fact that with increasing frequency k_h approaches $\omega/c(h)$, so that $\tan\theta(h, k_h)$ approaches zero and $f(k_h)$ becomes very large. We therefore have no assurance that the second term will be small for any frequency. This convergence difficulty arises from the fact that the asymptotic approximations used for the function $N(z, k)$

[7] See, for example, the treatment of the method of stationary phase given by C. Eckart, Revs. Modern Phys. **20**, 399 (1948).

are not valid for values of z and k such that $\tan\theta$ is very small. If it were possible to obtain exact solutions of the differential equation, the convergence difficulty would presumably not occur. We therefore take Eq. (9.06) as it stands, without any correction term, as representing the best approximation that can be found by the present method. As justification for this procedure we shall show that, except for the diffraction effects represented by the fresnel integral, Eq. (9.06) leads to exactly the same expression for the energy flux density that one derives on the basis of ray geometry.

When v_b and v_a are large in absolute value, the fresnel integral approaches the value $(2i)^{\frac{1}{2}}$, so that

$$I\rightarrow 2[irK(\partial R/\partial K)\tan\theta(z,K)\tan\theta(h,K)]^{-\frac{1}{2}} \times\exp\{-iK[r+X(K)]\}. \quad (9.09)$$

Calling $\phi(K)$ the contribution to ϕ that arises from any particular ray through the point (r, z), we have

$$\mathrm{Amp}\phi(K) = A[|RK(\partial R/\partial K)\tan\theta(z,K)\tan\theta(h,K)|]^{-\frac{1}{2}}. \quad (9.10)$$

In computing the particle velocity from ϕ we may treat Ampϕ as constant and differentiate only the exponential phase factor, since Ampϕ is, by comparison, a slowly varying quantity. Using Eq. (1.04) gives the absolute value of the energy flux density as

$$|J(K)| = [\rho(z)/c(z)][\omega\,\mathrm{Amp}\phi(K)]^2 = [\rho(z)\omega^2A^2/c(z)][|RK(\partial R/\partial K) \times\tan\theta(z,K)\tan\theta(h,K)|]^{-1}. \quad (9.11)$$

From the point of view of ray geometry, we consider[8] the energy flux emitted in the solid angle between the rays that leave the source at angles $\theta(h, K)$ and $\theta(h, K)+d\theta$ (see Fig. 11). If E is the total rate of energy emission at the source, we have

$$dE=(E/2)\cos\theta(h,K)d\theta=(E/2)d[\sin\theta(h,K)].$$

The area of wave front included between these rays at the point $[z, R(K)]$ is

$$dS=-2\pi R\sin\theta(z,K)dR.$$

Hence,

$$|J|=dE/dS = E[-4\pi R\sin\theta(z,K)dR/d\{\sin\theta(h,K)\}]^{-1}. \quad (9.12)$$

Now $dR/d[\sin\theta(h,K)]=(\partial R/\partial K)\{\partial[\sin\theta(h,K)]/\partial K\}$, and using

$$\sin\theta(h,K)=[1-K^2c^2(h)/\omega^2]^{\frac{1}{2}},$$

with

$$\cos\theta(z,K)/\cos\theta(h,K)=c(z)/c(h),$$

Eq. (9.12) may be written in the form

$$|J|=[Ec(h)/4\pi c(z)] \times[|RK(\partial R/\partial K)\tan\theta(z,K)\tan\theta(h,K)|]^{-1}. \quad (9.13)$$

[8] This treatment follows a similar derivation by B. Gutenberg, J. Acoust. Soc. Am. 14, 151 (1942), but is repeated here, since the identity between Gutenberg's equation and Eq. (9.11) is not obvious without some changes in notation.

From the limiting form approached by ϕ in the neighborhood of the point source, E and A are connected by the relation

$$E=4\pi\rho(h)\omega^2A^2/c(h), \quad (9.14)$$

so that Eqs. (9.13) and (9.11) are identical except that $\rho(z)$ appears in Eq. (9.11) in place of the $\rho(h)$ in Eq. (9.13). This discrepancy is removed by applying to Eq. (9.11) the density correction factor $\rho(h)/\rho(z)$ previously discussed.

Case II. (r, z) lies in a shadow zone where the boundary of the shadow zone is formed by a single ray (i.e., the boundary is not a caustic).

Let K be the parameter specifying the bounding ray. Then

$$I\simeq f(K)\int_{k_a}^{k_b}\exp\{-i[a+b(k-K)+c(k-K)^2]\}dk,$$

where $a=K[r+X(K)]$, $b=r-R(K)$, $c=-(\frac{1}{2})(\partial R/\partial K)$. Substitution of $v=[2/\pi c]^{\frac{1}{2}}[c(k-K)+b/2]$ as the variable of integration leads to

$$I\simeq[-\pi/(\partial R/\partial K)]^{\frac{1}{2}}f(K)\exp[-i\{K[r+X(K)] +[r-R(K)]^2/(2\partial R/\partial K)\}] \times\int_{v_a}^{v_b}\exp(-i\pi v^2/2)dv, \quad (9.15)$$

where

$$v_a=[-\pi\partial R/\partial K]^{-\frac{1}{2}}[r-R(K)-(k_a-K)\partial R/\partial K], \quad (9.16)$$
$$v_b=[-\pi\partial R/\partial K]^{-\frac{1}{2}}[r-R(K)-(k_b-K)\partial R/\partial K]. \quad (9.17)$$

Case III. (r, z) lies near a point such that $\partial R/\partial k=0$ is satisfied for some value of $k=K$. The appropriate expression for I in this case is

$$I\simeq[6/(\partial^2R/\partial K^2)]^{\frac{1}{3}}f(k)\exp\{-iK[r+X(K)]\} \times\int_{v_a}^{v_b}\exp[i(wv+v^3)]dv, \quad (9.18)$$

where

$$w=[R(K)-r][6/(\partial^2R/\partial K^2)]^{\frac{1}{3}}, \quad (9.19)$$
$$v_a=(k_a-K)[(\partial^2R/\partial K^2)/6]^{\frac{1}{3}}, \quad (9.20)$$
$$v_b=(k_b-K)[(\partial^2R/\partial K^2)/6]^{\frac{1}{3}}. \quad (9.21)$$

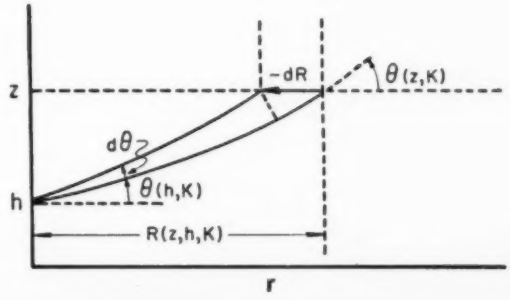


FIG. 11. Ray diagram near the source.

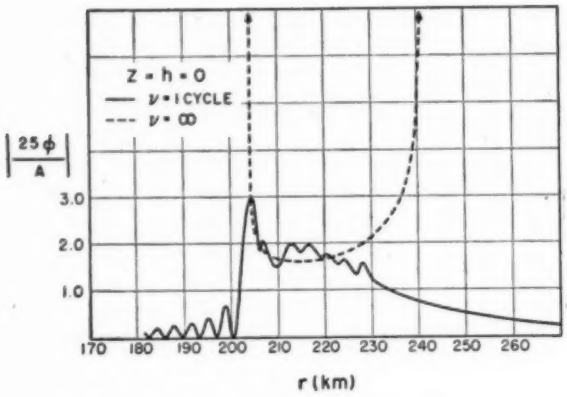


FIG. 12. Relative amplitude distribution in and near first geometric wave zone.

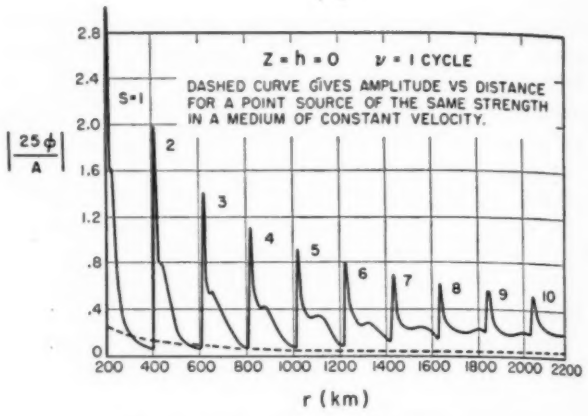

Z = h = 0 ν = 1 CYCLE
DASHED CURVE GIVES AMPLITUDE VS DISTANCE FOR A POINT SOURCE OF THE SAME STRENGTH IN A MEDIUM OF CONSTANT VELOCITY.
S=1
|25φ/A|
2.8
2.4
2.0
1.6
1.2
.8
.4
0
2 3 4 5 6 7 8 9 10
200 400 600 800 1000 1200 1400 1600 1800 2000 2200
r (km)

FIG. 13. Relative amplitude at long ranges.

When v_b and v_a are large in absolute magnitude, the integral in Eq. (9.18) approaches the airy integral (except for a constant factor),

$$\int_{-\infty}^{+\infty} \exp[i(wv+v^3)]dv$$

$$= \begin{cases} (2\pi/3)(|w|/3)^{\frac{1}{2}}\{I_{-\frac{1}{3}}[2(|w|/3)^{\frac{3}{2}}] \\ \qquad -I_{\frac{1}{3}}[2(|w|/3)^{\frac{3}{2}}]\} & w>0, \\ (2\pi/3)(|w|/3)^{\frac{1}{2}}\{J_{-\frac{1}{3}}[2(|w|/3)^{\frac{3}{2}}] \\ \qquad -J_{\frac{1}{3}}[2(|w|/3)^{\frac{3}{2}}]\} & w<0. \end{cases} \tag{9.22}$$

Numerical values of $|\varphi|$ as a function of r have been computed for a frequency of 1 cycle in the case of a source and receiver at the ground surface ($z=h=0$). With the velocity-altitude function tabulated above, rays from the source return to ground level within the range intervals 204.1 to 240.6 km (first geometric wave zone), 408.2 to 481.2 km (second geometric wave zone), etc. Values of $|\varphi|$ in an near the first geometric wave zone are plotted in Fig. 12. In this case the inner boundary lies on a caustic and the amplitude for values of r less than about 206 km is given by Eq. (9.18). For the particular numerical values used, the limits of integration, v_a and v_b, are not large in absolute magnitude (being 1.589 and -0.553, respectively), so that the airy integral with infinite limits is not a good approximation. When v_a and v_b are small in absolute magnitude an alternative approximation may be obtained by neglecting the v^3 term in the exponent. The values plotted in Fig. 12 for $r<206$ km have been computed from this alternative approximation.

Values of $|\varphi|$ at longer ranges, with the minor diffraction minima and maxima ignored, are shown in Fig. 13. In computing the height of the peak at the inner boundary of the successive geometric wave zones the airy integral has been used for the second and more distant zones, since the limits of integration increase with increasing values of s. Figure 13 illustrates the spreading of wave energy by diffraction into the shadow zones until, beyond the third geometric wave zone, the amplitude is every where greater than it would be for a point source of the same strength in a medium of constant velocity.

On the X-Ray Absorption Correction for Encased Diffracters in the Debye-Scherrer Technique*

H. L. Ritter, R. L. Harris, and R. E. Wood
Department of Chemistry, University of Wisconsin, Madison, Wisconsin
(Received June 23, 1950)

In dealing with numerical values for relative intensities derived from the Debye-Scherrer technique, the alteration of diffracted intensities produced by absorption in the sample cannot be ignored. Claassen has shown how this effect can be allowed for in the case of homogeneous cylindrical samples. This paper deals with a sample consisting of a cylinder surrounded by another annular cylinder both when the core is inert (i.e., absorbing and producing only extraneous scattering) and when the annulus is inert. The method is applied to show how the diffraction pattern for a liquid enclosed in a glass container may be corrected for the diffuse diffraction of the container as well as for absorption in the liquid and in the glass.

INTRODUCTION

When a diffracting specimen is irradiated by x-rays, the diffracted intensities are considerably altered by absorption in the sample of both the primary and the diffracted x-ray beams. Since the magnitude of this effect is angle-dependent, the relative diffracted intensities observed at different angles may be quite different from what they would have been in the absence of absorption. The necessity for allowing for this effect when numerical values of relative diffracted intensities are sought has been discussed by Blake.[1] A method for computing the self-absorption correction for circular cylindrical diffractors was applied by Debye and Scherrer[2] as early as 1918. Actual procedures have been described by Claassen,[3] Rusterholz,[4] Möller and Reis,[5] and extended by Blake. Tables of the resulting correction as a function of the diffraction angle, θ, and μR, the product of the absorption coefficient by the radius of the sample, are given by these authors and may also be found in the *Internationale Tabellen*.[6] The effect is further complicated when the diffracting specimen is either encased in an inert container or coated on the outside of an inert core. In these cases corrections should be made for absorption in the support as well. It is implicit in the tables quoted, however, and it is shown explicitly below, that the corrections here are essentially independent of diffraction angle providing the support is small in dimensions and of low absorption. Thus, the only effect in the case of a thin support is to produce an over-all and uniform lowering of intensity with no resulting alteration in relative intensities.

The effect is still further complicated, however, when both diffracting sample and support are noncrystalline, for example, when the sample is a liquid enclosed in an annular cylinder of glass. If one now seeks the diffracted intensity for the liquid sample as a function of diffraction angle, correction must be made both for the diffuse pattern due to the glass and for absorption in glass and sample. The correction for the diffuse diffraction pattern of the glass cannot be made simply by subtracting the pattern obtained by exposing independently the glass alone for an equivalent length of time. For the pattern so produced is affected by self-absorption in the empty annular cylinder with no absorption in the core whereas the contribution of the glass to the total pattern is modified by absorption in the core. The pattern for the support alone must thus also be corrected for absorption before subtracting out its contribution from the total. Finally, the remaining pattern, now due only to diffraction from the sample core, must be corrected for both self-absorption and absorption in the annular casing. This paper describes the methods used by the authors for making each of these corrections:

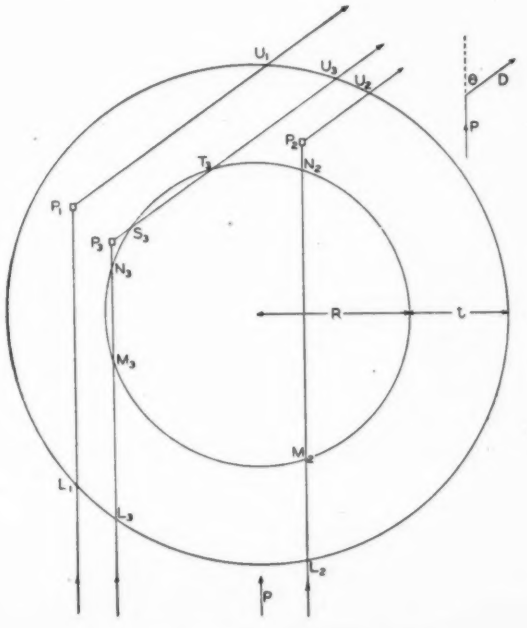


FIG. 1. Geometry of the correction $A_{a,a}$.

* Presented at the 1950 meeting of the American Crystallographic Association, State College, Pennsylvania, April 11, 1950.
[1] F. C. Blake, Revs. Modern Phys. **5**, 178 (1933).
[2] P. Debye and P. Scherrer, Physik. Z. **29**, 474 (1918).
[3] A. Claassen, Phil. Mag. **9**, 57 (1930).
[4] A. Rusterholz, Helv. Phys. Acta **4**, 68 (1931).
[5] H. Möller and A. Reis, Z. physik. Chem. **A139**, 425 (1929).
[6] *Internationale Tabellen Zur Bestimmung von Kristallstrukturen* (Berlin, 1935).

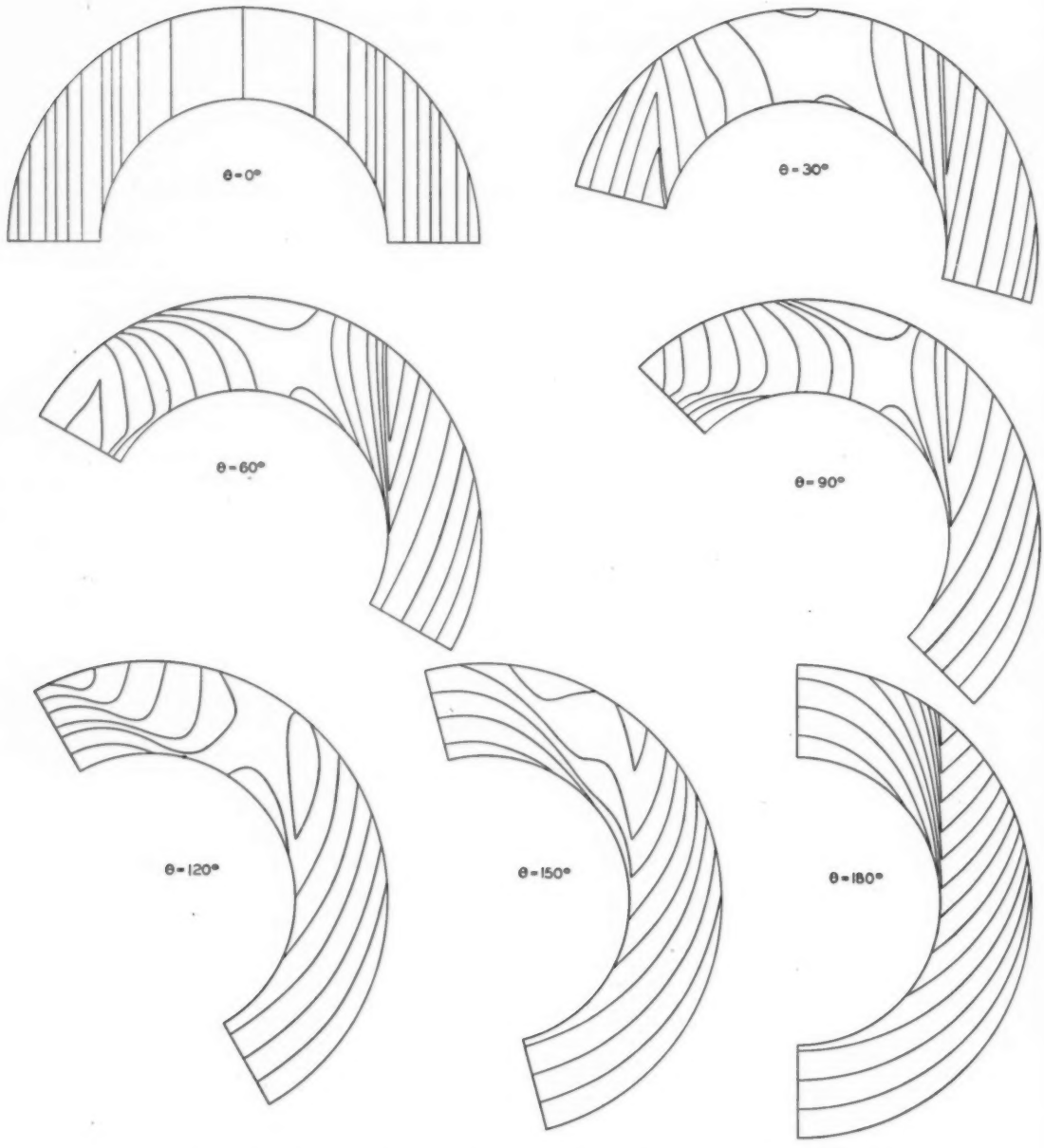


FIG. 2. Set of graphical integrands for $A_{a,a}$. Drawn for the case where $t/R=0.65$.

1. For diffraction in the annulus, absorption in the annulus ($A_{a,a}$).
2. For diffraction in the annulus, absorption in the annulus and core ($A_{a,ac}$).
3. For diffraction in the core, absorption in the annulus and core ($A_{c,ac}$).

The remaining case of diffraction in the core and self-absorption ($A_{c,c}$) is, of course, the case treated by Claassen.

SELF-ABSORPTION IN AN ANNULAR CYLINDER ($A_{a,a}$)

Consider a cross section of the cylinder lying in the equatorial plane (the plane containing the primary beam and normal to the cylinder axis), Fig. 1. The section is bathed by a primary beam in the direction of the arrow P, and we are concerned with diffraction taking place in the direction of the arrow D. Consider an element of diffracting area $d\omega_1$ at P_1 in the annulus. The beam (before and after diffraction) travels through the medium a distance $L_1P_1U_1=d_1$. Other diffracting areas $d\omega_2$, $d\omega_3$, $\cdots$ have associated with them other d's, which may be in two or three (as from P_2) or four (as from P_3) pieces. The intensity received on the film in the direction of D will be decreased by absorption by the factor $e^{-\mu d}$ for every elementary diffracting area $d\omega$. That is, the intensity dI contributed to the total intensity by diffraction from $d\omega$ will be proportional to $e^{-\mu d}d\omega$ rather than to $d\omega$ as would be the case with no absorption. The total intensity received by the

film in this direction will be proportional to $\int e^{-\mu d} d\omega$ rather than simply to $\int d\omega = \Omega$. We seek the value of $\int e^{-\mu d} d\omega/\Omega$ as a function of θ. The correction to be applied as a multiplicative factor to the observed intensities is then $1/A_{a,a}$, where $A_{a,a} = \int e^{-\mu d} d\omega/\Omega$.

This integral must be evaluated numerically. Following Claassen, we have effected the integration by dividing the annular area, Ω, into strips of essentially constant integrand, $e^{-\mu d_i}$, and replacing the integral by the summation $\sum e^{-\mu d_i} d\omega_i$. To lay out strips of (nearly) constant d_i, the area is divided by loci of equal d_i (equitraversals), the area bounded by two successive and sufficiently close equitraversals having a value of d_i about midway between the d_i's associated with the equitraversals themselves. Figure 2 shows a set of annuli broken into strips of about constant d_i in this way. The resulting integral is dependent upon θ, μR, and t/R.

The method for constructing the equitraversals may be seen by referring to Fig. 3. This figure represents the same plane section of the sample with the primary beam entering in the direction of the arrow. The curves drawn in over the annular area are loci of equal distances to the perimeter of the outer circle in a direction parallel to the arrow. They may be plotted most conveniently by measuring inward from the perimeter distances of $d = 0.1R, 0.2R, \cdots$, etc., with the help of some kind of parallel ruling device. For example the segments LP and the summed-segments LM and NP are all equal in length. This means that if a primary beam enters the sample in the direction indicated and travels before diffraction to any point P in the annulus, it will travel the same distance in the annulus before diffraction for every point P on any one locus. Associated with each locus then is a partial d representing this "incoming" partial traverse, δ. By reversing the arrow, it is further clear that if a *diffracted* ("outgoing") beam originates at any point P, it will travel the same distance (the same δ) through the annulus in the direction of the *reversed* arrow before *leaving* the sample, for all P's on any one locus.

Using now a drawing of Fig. 3 (hereafter called the opaque) and a tracing of the same figure on transparent drawing cloth (hereafter called the mask) *but with the arrow reversed*, the construction of the required equitraversals is simple. The mask is placed over the opaque with their centers coincident and with the two arrows oriented in such a way that one of them represents the direction of the primary beam and the other the direction of the diffracted beam. The intersection of any two loci, one on the opaque, δ_0, and one on the mask, δ_m, now represents a point P in the annulus such that the distance (δ_0) the primary beam travels *to* the point and the distance (δ_m) the diffracted beam travels *from* the point equals $\delta_0 + \delta_m = d$, all distances being measured in the annulus only. Other intersections may be found such that the sum $\delta_0 + \delta_m$ equals the same d. All the intersections for a fixed d may conveniently be plotted directly on another blank piece of tracing cloth superposed over the opaque-mask assembly, and then connected to form the required equitraversal. By selecting other convenient values of d the entire annular area is readily divided into small strips of area $d\omega_i$ over which $e^{-\mu d_i}$ is essentially constant. There is reproduced in Fig. 2 the set of so-divided annuli for seven values of θ from 0° to 180°. The diagrams are symmetrical about the line bisecting the angle between the primary and diffracted beams; only one of each symmetrical half is shown. It is worth noting that only one diagram like Fig. 3 is necessary for any combination of the parameters θ and μR which enter the function $A_{a,a}$. For any given μ, one diagram of equitraversals like those of Fig. 2 will serve for all μR. New Figs. 2 and 3 must be constructed when the experimental arrangement uses a new value of t/R.

The actual integration now involves evaluating the sum $\sum e^{-\mu d_i} d\omega_i$ over the area of the annulus. The authors effected this summation by making a contact print on a good grade of photographic paper of the annulus and its equitraversals as plotted directly on the tracing cloth and cutting out and weighing the elementary areas. Since the ratio of the weight of a strip to the weight of the entire annulus is $d\omega_i/\Omega$, the sum $\sum e^{-\mu d_i} d\omega_i/\Omega$ is directly $A_{a,a}$.

In order to illustrate the actual numerical work, some of the pertinent computed values are summarized in Table I. The case represented is that for which $t/R = 0.65$ and $\theta = 180°$. The corresponding equitraversal diagram is shown in Fig. 2. The annulus there is shown divided into thirteen elementary areas by means of equitraversals corresponding to $d/R = 0.0, 0.4, 0.8, \cdots 5.2$. In Table I are given the values of the d/R for each elementary area (taken as the mean of the d/R's for the bounding equitraversals) and the relative weights $d\omega/\Omega$

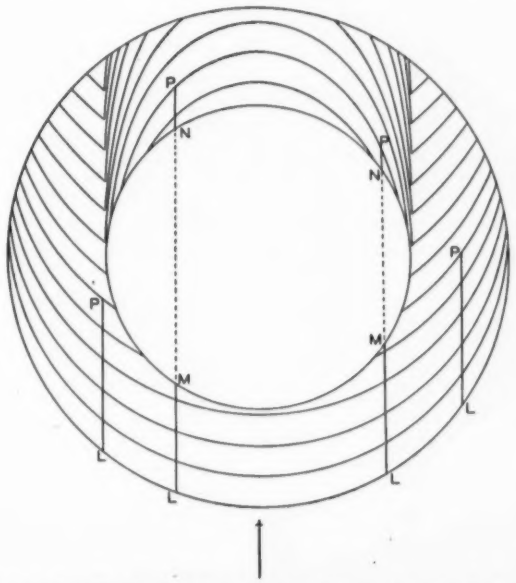


FIG. 3 Auxiliary mask for constructing equitraversals (a, a).

TABLE I. Sample calculation of $A_{a,a}$ for two typical cases.

d/R	$d\omega/\Omega$	Value of $e^{-\mu d}d\omega/\Omega$ for $\mu R=0.166$	$\mu R=1.55$
0.2	0.1213	0.1173	0.0889
0.6	0.1174	0.1062	0.0464
1.0	0.1213	0.1027	0.0257
1.4	0.1188	0.0942	0.0135
1.8	0.1173	0.0869	0.0072
2.2	0.1192	0.0827	0.0039
2.6	0.0971	0.0632	0.0017
3.0	0.0619	0.0376	0.0006
3.4	0.0460	0.0261	0.0002
3.8	0.0346	0.0184	0.0001
4.2	0.0249	0.0124	
4.6	0.0142	0.0066	
5.0	0.0058	0.0026	

of these areas. Three of the elementary areas are in two pieces in the figure. Their weights have been combined in the table. Also tabulated are the values of $e^{-\mu d}d\omega/\Omega$ for the two cases $\mu R=0.166$ and $\mu R=1.55$. These values for μR correspond to, say, absorption of MoK_α and CuK_α radiation, respectively, in a glass tube of 0.21-mm inner radius. The sums of the several products $e^{-\mu d}d\omega/\Omega$ are directly the values of $A_{a,a}$.

To show the dependence of $A_{a,a}$ on θ for different combinations of the parameters μR and t/R, Table II gives the value of $A_{a,a}(\theta)$ for some selected combinations of high and low μR and high and low t/R. It is interesting to note that the alteration in intensity is considerable (12 percent) even in the case of thin tubing ($t/R=0.25$) and low absorption ($\mu R=0.160$). But it should be further noted that even when the value of $A_{a,a}$ departs appreciably from unity, it may still be nearly enough independent of θ that *relative* intensities are not much changed over the entire angular range. By using the procedure described above, the entire function $A_{a,a}(\theta)$ for one pair of parameters t/R and μR—including drawing the mask and opaque, plotting the equitraversals for seven values of θ, cutting out and weighing the $d\omega$'s, and evaluating the sums $\sum e^{-\mu d}d\omega/\Omega$—may be evaluated in about eight man-hours. Normally, the full range of θ from 0 to π will not be required.

ABSORPTION IN ANNULAR CYLINDER AND CORE FOR DIFFRACTION IN THE ANNULUS ($A_{a,ac}$)

Figure 4 pictures again the same plane section of the annulus this time filled with an absorbing material.

TABLE II. Dependence of $A_{a,a}(\theta)$ on μR and t/R.

R	0.166	1.55	0.160	1.60
t/R	0.65	0.65	0.25	0.25
	$A_{a,a}$	$A_{a,a}$	$A_{a,a}$	$A_{a,a}$
0	0.747	0.081	0.883	0.308
30°	0.748	0.091	0.883	0.310
60°	0.751	0.112	0.883	0.315
90°	0.753	0.133	0.883	0.325
120°	0.754	0.154	0.883	0.340
150°	0.756	0.176	0.883	0.358
180°	0.757	0.188	0.883	0.380

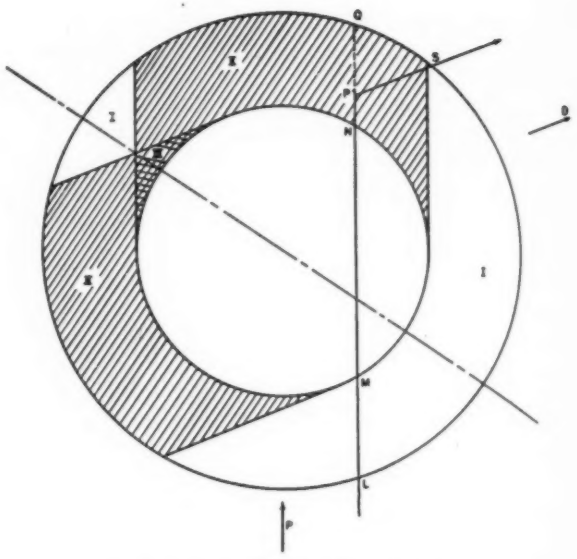


FIG. 4. Geometry of the correction $A_{a,ac}$.

Consider diffraction taking place only in the annulus, with the primary and diffracted beams directed as shown by the arrows. It is clear that the entire annular area may be divided into three distinct regions: In region I lie those diffracting centers for which both the primary beam and diffracted beam travel only through the annulus. In region II *either* the primary beam or the diffracted beam travels in part through the core. Because of the line of symmetry it is not necessary to consider both parts of this region. In region III *both* the primary and diffracted beams travel in part through the core. Because of the compound nature of the path, the integrand of the absorption integral must be replaced by $\mu_c d_c+\mu_a d_a$, where d_c and d_a are, respectively, the distances traveled in the core and in the annulus, and μ_c and μ_a are the corresponding absorption coefficients. Because two different absorption coefficients are now involved, the plotting of equitraversals is no longer significant. We seek now to divide the annular area into elementary areas of essentially constant absorption, $\mu_c d_c+\mu_a d_a$. In region I (the back segment of this region exists only when $\theta<2$ arc cos $t/(R+t)$), $d_c=0$ and the lines of constant absorption are the same as the equitraversals of the case (a, c). In region II there are contributions to the total μd from both the core and the annulus. To plot lines of equal absorption in this region, it is convenient to have at hand (1) a set of drawings showing the equitraversals of the case (a, a) and (2) a set of drawings showing equitraversals of the core for diffraction in the annulus (a, c). The former of these (a, a) show lines of constant traverse through the annulus; i.e., lines of constant $\mu_a d_a$. The latter (a, c) show lines of constant $\mu_c d_c$. By superposing the one on the other in an opaque-mask assembly such as previously described it is a simple matter to plot on a third sheet of blank drawing cloth the loci of constant $\mu_a d_a+\mu_c d_c$. A method for obtaining the masks (a, a) was

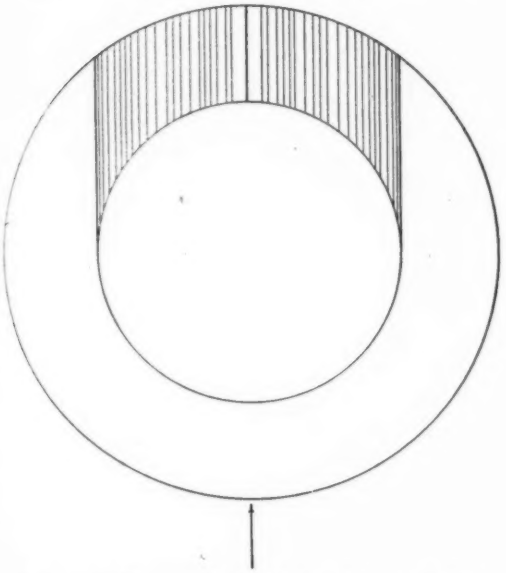


FIG. 5. Auxiliary mask showing equitraversals (a, c).

described above; a method for sketching the masks (a, c) follows.

Referring again to Fig. 4, consider the beam which undergoes diffraction at P. This beam has traveled a distance $d_a = LM + NP + PS$ in the annulus and a distance $d_c = MN$ in the core. Any point on the straight line NPQ will have associated with it a $d_c = \mathrm{MN}$; and NPQ is therefore an equitraversal (a, c) whose actual traverse in the core is MN. Figure 5 shows a plot of this mask which is valid for all θ. It is simplest to construct this mask by assigning to each of the parallel lines a d_c calculated from its distance from the parallel diameter.

In use, lines of constant total μd for a chosen θ are constructed by superposing the mask of Fig. 5 on the proper member of the set in Fig. 2, with the arrows representing the primary beam both pointing in the same direction. With the help of the equitraversals representing d_c and d_a now superposed, one now chooses a value for $\mu_a d_a + \mu_c d_c$ and plots combinations (of d_c and d_a which yield the required total absorption) on a third sheet of blank tracing cloth. By choosing a series of sufficiently closely spaced values for the total absorption, the entire annulus may be filled in with loci. In region I only the mask (a, a) will show any equitraversals, and in this region $\mu_a d_a$ is plotted directly. In region II both masks show lines and both (a, a) and (a, c) contribute to the total μd. For region III it is helpful to have still a third mask $(a, c)'$, which is exactly like (a, c) except with the arrow reversed. This mask is oriented with its arrow pointing in the direction of the diffracted beam, and then all three masks contribute to the total μd. [It is perhaps worth remarking that in actual practice short cuts of this procedure inevitably occur to the operator which are a great deal easier to devise than to describe.]

TABLE III. Computed values of absorption corrections for a typical case: $t/R = 0.65$, $R = 0.021$ cm, $\mu_a = 7.9$, $\mu_c = 13.4$ cm^{-1}.

θ	$A_{a,a}$	$A_{a,ac}$	$A_{a,ac}/A_{a,a}$	$A_{c,ac}$	$A_{c,a}$*
0°	0.747	0.599	0.801	0.488	0.630
30°	0.748	0.601	0.802	0.490	0.630
60°	0.751	0.610	0.814	0.494	0.631
90°	0.753	0.621	0.824	0.498	0.632
120°	0.754	0.634	0.842	0.501	0.637
150°	0.756	0.648	0.858	0.504	0.643
180°	0.757	0.655	0.865	0.508	0.647

* From the *Internationale Tabellen*.

Figure 6 shows some typical resultant plots of lines of equal absorption for the case in which $t/R = 0.65$ and μ_a and μ_c are, respectively, 7.9 and 13.4 cm^{-1}. The plotted curves are lines of constant $(\mu_a d_a + \mu_c d_c)/R$. The integration is carried out in the same way as previously described to yield the value of $A_{a,ac}$. Results for a typical case are given in Table III. In order to correct for the contribution to the total pattern of the tubular container alone, the intensities observed in an equivalent exposure of the container alone are multiplied by the ratio $A_{a,ac}/A_{a,a}$ and the resulting corrected intensities subtracted from the total observed intensities.

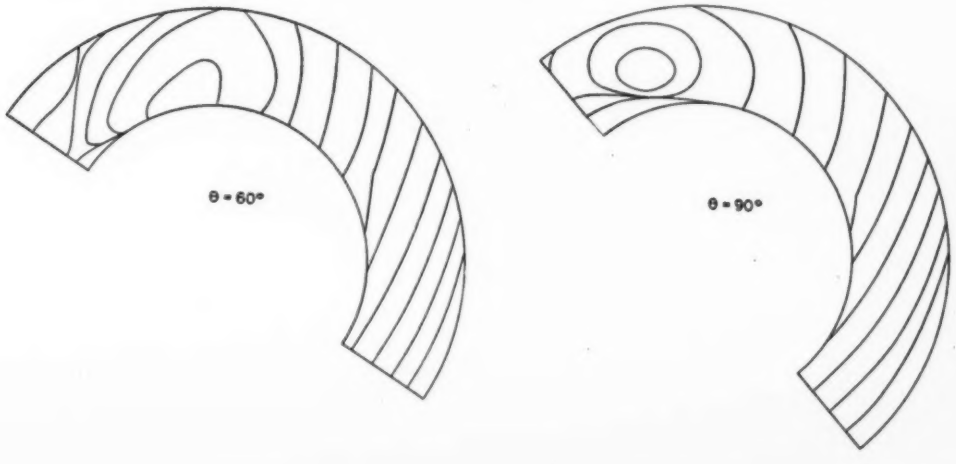


FIG. 6. Typical graphical integrands for $A_{a,ac}$. Drawn for the case where $t/R = 0.65$, $\mu_a = 7.9$ cm^{-1}, and $\mu_c = 13.4$ cm^{-1}.

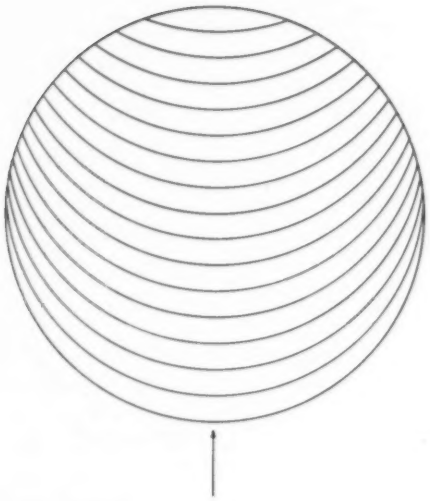


FIG. 7. Auxiliary mask for constructing equitraversals (c, c).

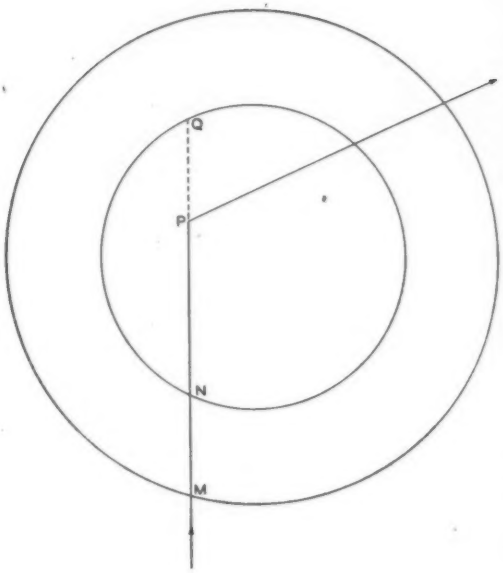


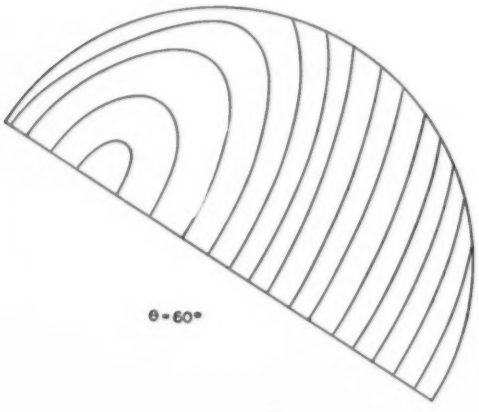
FIG. 9. Geometry involved in equitraversals (c, a).

ABSORPTION IN ANNULAR CYLINDER AND CORE FOR DIFFRACTION IN THE CORE ($A_{c,ac}$)

In this case again the integrand is composed of two parts because the beam path passes through both media. As in the preceding case, it is convenient to construct the lines of constant absorption by superposing masks showing equitraversals (c, c) and (c, a). Equitraversals of the type (c, c) may be constructed as discussed by Claassen; the following adaptation of his construction is simple to carry out.

Figure 7 represents the equatorial cross section of the core alone, radius R. The curves drawn across the face of the section are arcs of circles. They have the property that the distance, δ, parallel to the direction of the primary beam from any point on one of the curves to the perimeter of the circle R is constant. By superposing two such drawings, one as shown (the opaque) and the other differing only in having the arrow reversed (the mask), such that the arrow of the opaque and that of the mask are oriented respectively as the primary and diffracted beams; then a series of intersections may again be plotted on a third transparent sheet such that for each intersection $\delta_0+\delta_m=d_c$. By plotting a set of equitraversals across the face of the section in this way one obtains a drawing like those of Fig. 8.

The equitraversals (c, a) are obtained similarly. The geometry involved is shown in Fig. 9, representing the annular cross section. Consider only the path traversed by the primary beam before diffraction at P. No matter where on the chord QN the point P lies, the distance traversed by the beam through the annulus is the same distance, MN. QN is therefore the locus of all diffracting points having the property that the primary beam traverses the same thickness, δ, of annulus before diffraction. Figure 10 shows a diagram of the whole set of such loci for the case where $t/R=0.65$. The distance, x, of any of the parallel chords from the primary diameter can easily be shown to be given by $4\delta^2x^2=R^2[(2+t/R)^2-\delta^2][\delta^2-(t/R)^2]$. One can make a duplicate of this drawing on tracing cloth but with the arrow reversed, and by plotting loci of constant $d_a=\delta_0+\delta_m$ from an opaque-mask assembly as described above, a new mask



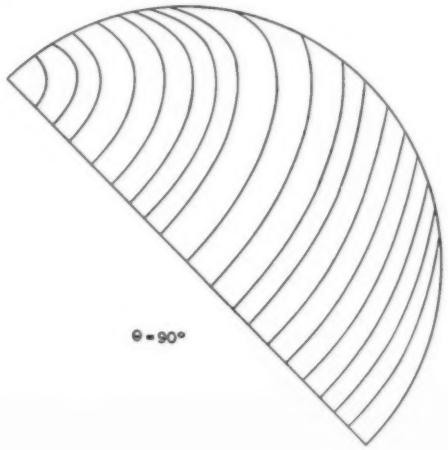


FIG. 8. Typical masks showing equitraversals (c, c).

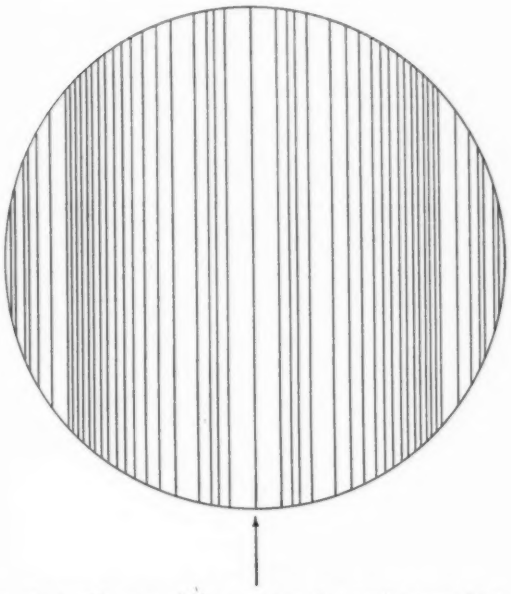


FIG. 10. Auxiliary mask for constructing equitraversals (c, a).

may be prepared similar to those of Fig. 11, typical equitraversals (c, a) for the case $t/R=0.65$.

With the help of the pair of equitraversal masks (c, c) and (c, a) for a chosen θ, one can now superpose them and plot lines of constant absorption, $\mu d=\mu_a d_a+\mu_c d_c$, exactly as described for $A_{a,ac}$. Figure 12 shows a typical plot of the lines of constant absorption for the same values of the constants as in Fig. 6. Integration over this area yields $A_{c,ac}$. Table III shows the corresponding values of the absorption correction.

APPLICATION OF THE CORRECTIONS

Two of the standard procedures in powder practice use cylindrical samples. In one, the diffracting sample is placed within an inert tubular casing; in the other, the diffracting sample is coated around an inert cylindrical core. Accurate values for the relative diffracted intensities must take into consideration the modification of these intensities not only by self-absorption but also by absorption in the support as well. When it is necessary to make this correction, the values of $A_{a,ac}$ (for the sample coated on an inert rod) and of $A_{c,ac}$ (for the sample enclosed in a tube) may be computed as outlined in the preceding discussion and applied. The observed intensities are converted to absorption-corrected intensities by dividing by the factor A. It should be noted, however, that with the current practice of using thin glass supports, the correction will rarely be necessary. For example, consider the case of aluminum chlo-

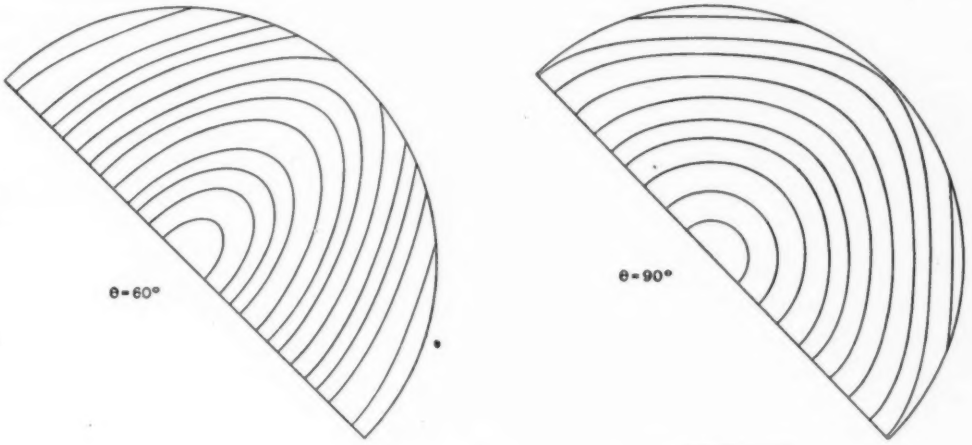


FIG. 11. Typical masks showing equitraversals (c, a).

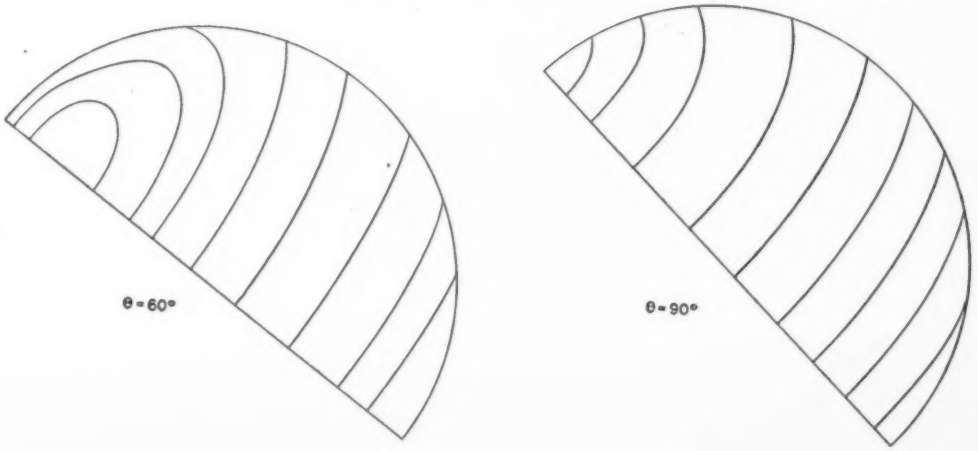


FIG. 12. Typical graphical integrands for $A_{c,ac}$. Drawn for the same parameters as in Fig. 6.

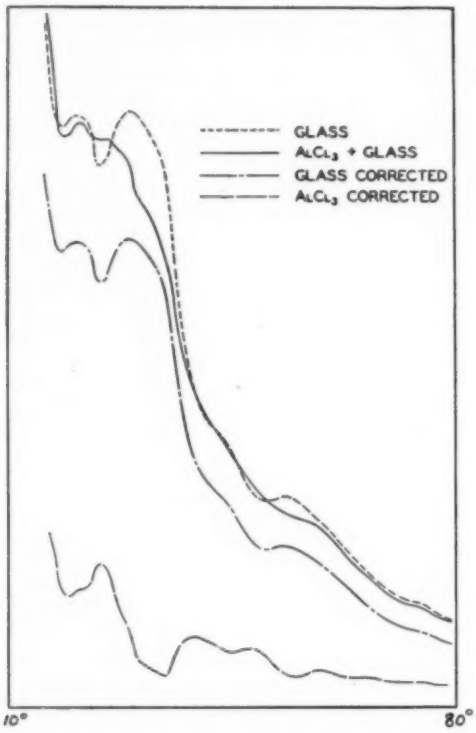


FIG. 13. Diffraction patterns for liquid aluminum chloride. Angle of diffraction, degrees. Diffracted intensity in arbitrary units.

ride contained in a glass tube of dimensions as given in Table III, MoK$_\alpha$ radiation. The column $A_{c,ac}$ gives the combined correction for this case and it will be noted that the absorption correction scarcely varies by as much as 5 percent. Provided that relative intensities are sought with an accuracy within this range, the correction may be omitted. More important, however, is the observation that the self-absorption corrections, $A_{c,c}$ alone, available from the *Internationale Tabellen* without laborious computation, are very nearly proportional (within less than 2 percent) to the complete correction. The deviation from proportionality will be considerably less if thinner glass ($t/R=0.65$ is very thick) is used and for materials of higher self-absorption than $\mu=13.4$ cm^{-1}. Thus, it appears that even when the necessity for making the correction is indicated, nearly all cases of powder diffraction may be satisfactorily handled by applying the correction $A_{c,c}$ from already published tables.

Similar considerations apply to the case of a sample coated on a thin glass fiber. The correction may be ignored completely; the correction may be applied partially by computing $A_{a,a}$ alone; or it may be applied completely by computing $A_{a,ac}$. The partial correction will again be very nearly proportional to the total correction, so that when only relative intensities are sought, the correction $A_{a,a}$ will almost always be sufficient. Even here it is possible to achieve a good approximation to $A_{a,a}$ by entering the usual tables for $A_{c,c}$ with an "equivalent" value of R (the *Internationale Tabellen*'s R) somewhat smaller than the outer radius of the sample.

The greatest utility of the corrections discussed here lies in their application to the diffraction patterns of liquids enclosed in glass tubes. The solid curve of Fig. 13 shows a plot of the microphotometer trace of the diffraction pattern of liquid $AlCl_3$ in a glass tube, and the dotted curve shows that for the tube alone, both taken under the same total x-ray flux. In order to subtract from the total pattern the contribution made by the glass alone, each ordinate of the dotted curve was multiplied by $A_{a,ac}/A_{a,a}$. The resulting function is plotted as the "glass corrected" curve. The difference between these two curves is plotted at the bottom of Fig. 13, and represents the raw diffraction pattern for liquid $AlCl_3$ alone. After division of these ordinates by the factor $A_{c,ac}$ there results the diffraction pattern for liquid $AlCl_3$ fully corrected for extraneous diffraction, self-absorption, and absorption in the container.

ACKNOWLEDGMENT

The authors wish to record their thanks to the Wisconsin Alumni Research Foundation and to the Du Pont Company for grants which facilitated their work.

Dry Metallic Friction as a Function of Temperature Between 4.2°K and 600°K

I. Šimon, H. O. McMahon, and R. J. Bowen
Arthur D. Little, Inc., Cambridge, Massachusetts*
(Received September 8, 1950)

Coefficients of static friction of six metals (Fe, Ni, Cu, Pb, Sn, and Zn) have been measured at temperatures ranging from liquid helium up to room conditions (and in two cases above this). Shear strengths and hardnesses of these metals were measured over the same range of temperatures. It was found that in spite of the generally large variation of both of these latter quantities with temperature, their ratio is in most cases dependent, to only a slight extent, on temperature, following rather closely the change of the static-friction coefficient with temperature. This fact tends to substantiate the adhesion theory of friction proposed by F. P. Bowden.

INTRODUCTION

DESPITE the extensive literature concerning friction phenomena in general (for references see, e.g., Bowden and Tabor,[1] Gemant,[2] Bikerman[3]), only one paper, by Schnurmann,[4] appears to have been published on metallic friction at low temperatures (down to 78°K). The present investigation, which was limited to the problem of the friction between two unlubricated surfaces of the same metal, extended down to the temperature of liquid helium. Measurements of both the friction coefficients and mechanical properties of the metals over the same range of temperatures were obtained, and from these observations it is possible to make some evaluation of the nature of the elementary phenomena of dry friction, a subject which still appears to be open to various interpretations.[3]

In the surface-roughness hypothesis of Coulomb, it is assumed that the work of friction is expended in lifting the surface irregularities of one of the sliding members over those of the other. The coefficient of friction is then given by an expression of the form

$$\mu_r = \tan\vartheta,$$

where ϑ is the mean value of the slopes of the individual microscopic asperities. Since the surface profile is essentially independent of temperature, the coefficient of friction, μ_r, should not change with temperature. It has been shown, however, by Strang and Lewis[5] that the roughness component can account for only a few percent of the total observed friction.

According to the localized adhesion hypothesis of Bowden, the metal is subjected to such high pressures at the points of actual contact that it flows in a plastic manner. During this process, an adhesion or cold welding occurs. The shear resistance of any individual adhesion spot to an applied tangential force is then proportional to its cross section and to the shear strength of the welded metal. Since the cross section of actual contact is inversely proportional to the hardness of the metal and directly to the normal force, the coefficient of friction is

$$\mu_a = S/H,$$

where S and H denote the shear strength and hardness, respectively. This being true, μ_a will be a function of temperature, since both S and H show an appreciable variation with temperature.

Bowden also considered the friction work which is done by the tangential force in displacing metal out of the way of the sliding member (ploughing action). It appears to be much more difficult to set up an analytical expression for this phenomenon, except for special cases such as a hard material in contact with a very soft one. The ploughing portion of friction is probably influenced by temperature, which in turn changes mechanical properties of the metal. However, the contribution of ploughing to total observed friction has been found to be very small, even in the extreme instance just mentioned.[1]

The observed coefficient of friction, μ_s, is probably the resultant of all the actions noted above, modified by the strong "lubricating" influence of chemically bound surface films. However, if the slope of the μ_s *vs* T curve should coincide with that of the (S/H) *vs* T curve, the hypothesis of the predominance of the adhesion term would tend to be substantiated.

EXPERIMENTAL TECHNIQUES

1. Friction Apparatus

For determination of the static-friction coefficients at several temperatures between 4.2°K and 600°K, an apparatus was designed using the principle of a slider on an inclined plane. The static coefficients were obtained from the measured angles of repose by means of the formula, $\mu_s = \tan\theta$.

The general arrangement of the apparatus is indicated in Fig. 1. The friction samples (tube 5 and slider 6 in Fig. 1) were enclosed in a triple-walled glass vessel, 35 in. long and 3.75 in. o.d., which was pivoted about its lower end and could be tilted at any angle between the horizontal and the vertical. The purpose of the coaxial jacket formed by the two outermost walls was to provide a cooled shield (kept at the temperature of

* The work described in this paper was sponsored by the ONR under contract No. N-6-onr-22804.

[1] Bowden, Moore, and Tabor, J. Appl. Phys. **14**, 80 (1943).
[2] A. Gemant, J. Appl. Phys. **14**, 456 (1943).
[3] J. J. Bikerman, Revs. Modern Phys. **16**, 53 (1944).
[4] R. Schnurmann, Proc. Phys. Soc. (London) **53**, 538 (1941).
[5] C. D. Strang and C. R. Lewis, J. Appl. Phys. **20**, 1164 (1949).

liquid nitrogen) to reduce the heat influx during measurements in which liquid helium was the cooling bath for the sample. Other temperatures were obtained by cooling the tubular sample with liquid N_2 or solid-CO_2-acetone mixture, or heating it by means of an electric resistor inserted into the sample tube.

The glass enclosure was evacuated by means of a three-stage oil diffusion pump backed up by a two-stage rotary pump. A liquid-air trap was located ahead of the diffusion pump. The pressure was kept below 10^{-5} mm of mercury, and was measured by means of an ionization gauge.

A retaining ring on the end of a long wire inside the innermost glass tube was provided to allow the slider to be pulled back to the upper starting position after it had completed a slide. A permanent magnet manipulated from outside was used to move the wire by means of the iron ring attached to its upper end.

This arrangement was suitable for metals which were available in the form of tubing (usually 0.375 in. o.d.). In other cases the samples were made in the form of a V-shaped trough and a cylindrical slider and enclosed in a glass or fused silica tube (Fig. 2). Cooling was effected in such cases by immersing the tube containing the experiment in a suitable bath, such as liquid nitrogen or liquid helium. Helium gas was sealed into the tube to serve as a heat-transfer medium in order to speed cooling of the sample. The filling pressure at room temperature was 200 mm mercury. The angle of repose was measured by tilting the sample tube together with the Dewar vessels on a suitable support.

The friction data were taken by increasing the angle of tilt slowly until, under repeated light tapping on the glass enclosure, the slider just began to move downwards. The motion proceeded usually in a series of "slips and sticks," the velocity between two stops never exceeding a rather small value (of the order of 0.1 cm per sec). When this state was reached the tilting was stopped and the corresponding angle θ was recorded.

The coefficient of friction obtained in this way is actually not a true static coefficient but rather a kinetic one, corresponding to a very small average velocity.

In spite of the simple techniques used, the reproducibilities of the static data were satisfactory, the average deviation of repeated determinations amounting to about ± 5 percent. Data taken on different samples of the same metal, or on the same samples but during different runs, varied sometimes considerably more, a fact which was attributed to the impossibility of exactly duplicating the state of the surface.

2. Description of the Samples

Commercial grades of pure metals were used for most of the experiments. The only alloys used were a 0.16 percent carbon steel (both in normal state and case hardened) and an 18–8 chromium-nickel stainless steel (in normal state and nitrided). All metals used were polycrystalline.

For the experiments in sealed-off glass tubes (Fig. 2), the trough was made of sheet metal (0.020 to 0.032 in. thick, 8.5 in. long, 0.70 in. wide, angle of V: 90° or 60°). The slider was a cylinder, 1.0 in. long, 0.31 in. o.d., either cast and machined from the same specimen of metal or bent to a tubular shape from the sheet.

For the vacuum friction-apparatus, the tubular sample was made of a commercial tubing 0.375 in. o.d., closed at one end by welding or silver-soldering. The slider had the form of a tube 2 in. long, 0.50 or 0.625 in. o.d. with rounded edges. Sliders of other forms, such as inverted V-shaped troughs or crossed cylinders, were used occasionally. A description of the samples used and some of their mechanical properties can be found in Table I. The column, "Treatment," in this table refers to the initial state of the samples. Most of the samples were heated in vacuum up to 250°C to 300°C during the outgassing procedure. The hardness figures marked "B" are Brinell numbers obtained with an indenter described in the following section under conditions corresponding to the 500 kgm—10 mm—ball test. The figures marked "Rc" are Rockwell C-numbers obtained with a standard diamond penetrator. All measurements of these data were made at room temperature.

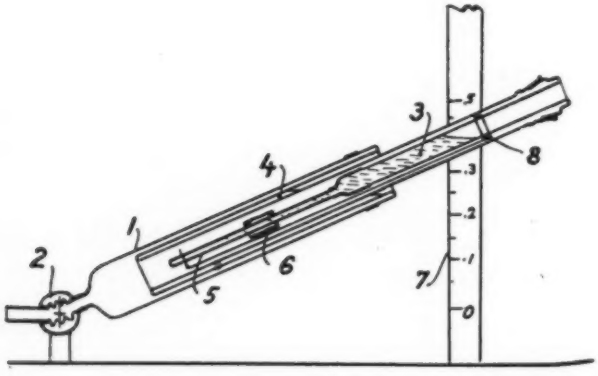


FIG. 1. Friction apparatus for interchangeable samples. (1—vacuum vessel, 2—pivot and flexible bellows, 3—liquid helium, 4—liquid-nitrogen jacket, 5—tubular sample, 6—slider, 7—tangent scale, 8—iron ring.)

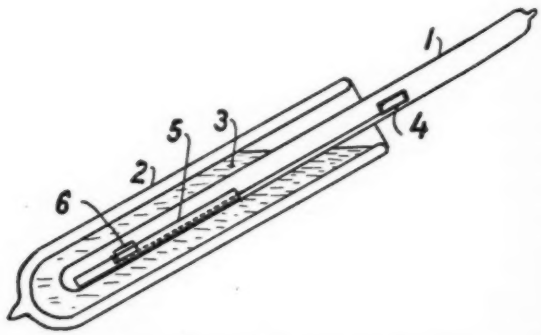


FIG. 2. Friction apparatus with enclosed sample. (1—sealed glass tube, 2—Dewar vessel, 3—liquid helium, 4—iron block for moving the slider, 5—trough-shaped sample, 6—slider.)

3. Measurement of Hardness and Shear Strength as Functions of Temperature

The Brinell-hardness was measured by means of an apparatus as shown in Fig. 3. A steel ball of 0.094 in. diameter is forced by means of a lever device into the surface of the sample by a load of 28.4 kgm, which corresponds to the conditions of the standard 500 kg-10-mm-ball test. The indenter can be immersed in a suitable cooling bath (including liquid helium) and the force applied when the temperature of the indenter, sample, and the bath have equalized. The diameters of the spherical impressions were measured under a microscope after warming-up, and the hardness numbers for the subsequent analysis were expressed as the ratio of the load to the projected area of the impression (Mayer's hardness number). The effect of thermal expansion during warming up was considered negligible. The effect of elastic recovery was also neglected, in accordance with the theory of contact hardness measurements developed by Holm and co-workers.[6] Since the indentation force, ball diameter, and contact time were held constant for these measurements, it seemed to be a justifiable assumption that the contact hardness at any one of the temperatures encountered would exhibit the same independence of the area of indentation found in the case of the room temperature determinations just referred to. Thus, the hardnesses were felt to be comparable.

No data could be taken on hardened and nitrided steels at low temperatures with this apparatus.

Regarding the shear strength of metals at low temperatures, the most feasible way to obtain reproducible data turned out to be a method of measuring the force necessary to shear a strip of metal in a punch and die apparatus, (Fig. 4). The actual shearing unit, shown in the figure, was used with the same loading device and in the same way as the hardness tester already described, (Fig. 3). Even though all the parts were made of hardened steel, its use was limited to the softer metals. It was necessary to keep the cross section of the sheared strips rather small, as an appreciable increase of shear strength occurs on cooling. A typical cross section was of the order of 0.40×0.25 mm^2 for the metals of the iron, steel, and nickel group, and 0.50×0.30 to 1.70×1.20 for the copper, zinc, and tin group. The cross section of the sheared area was measured under a microscope after completion of the experiment.

TABLE I. Description of the friction samples. (First line in each section pertains to the base, the second to the slider.)

Metal	Form	Treatment	Hardness	Av. grain size mm	Principal impurities (percent)
				mm	
Fe	Sheet	annealed in H_2 at	87B	0.82	0.08C
	Hot rolled rod	900°C, 3 hrs.	—	1.1	0.08C
Steel	Cold drawn tube	none	144B	—	0.16C
	Same	none	same	—	0.16C
Steel	Same	case hardened	C49 Rc	—	no analysis made
Stainless steel	Drawn tube	factory	137B	—	18 Cr, 8 Ni, 2 Mn, 1 Si, 0.15C
	Same	annealed	same	—	(Nominal composition)
Same	Same	nitrided	C48 Rc	—	no analysis made
Ni, (No. 1)	Drawn tube	factory annealed	120B	0.30	0.13 Fe, 0.19 Mn, 0.66 Co, 0.06C
	Sheet, hard	none	116B	—	0.21 Fe, 0.26 Mn, 0.36 Co, 0.05C
Ni, (No. 2)	Sheet, hard	none	116B	—	same
	Same	none	same	—	same
Cr	Electroplated on tubes	none	—	—	no analysis made
Cu	Soft tubing	factory	62B	0.62	0.015 P, 0.015 Sn, 0.007 Pb, 0.005 Fe
	Same	annealed	same	same	Mn, Si, O, absent
Pb, (No. 1)	Sheet	none	5.5B	—	0.019 Cu, 0.028 Bi, 0.003 Ag
	Cast in vacuum	none	—	—	less than 0.001 Sn
Pb (No. 2)	Brass tube, hot dipped in Pb	none	—	4.8	no analysis made
	Cast in vacuum	none	—	3	same as Pb No. 1
Sn	Sheet	none	7.9B	—	C.P. metal, no analysis made
	Cast in vacuum	none	—	—	
Zn	Galvanized sheet	none	—	5.3	no analysis made, C.P. metal, same
	Cast in vacuum	none	32B	2	

[6] Holm, Holm, and Shobert II, J. Appl. Phys. **20**, 319 (1949).

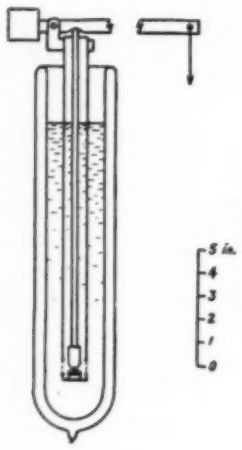


FIG. 3. Brinell-type indenter for measurement of hardness at low temperatures.

Due to uncertainty concerning the actual distribution of the stress in the sheared region, this method is not suitable for ascertaining the absolute value of shear strength of the metal under test. However, there are

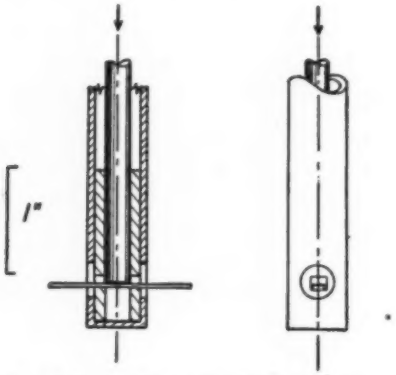


FIG. 4. Punch and die for shearing strips at low temperatures.

reasons for believing that the relative change of the measured force as a function of temperature corresponds to the same change in shear strength, at least for measurements taken on the same strip of any particular metal. Even with different strips of the same metal the

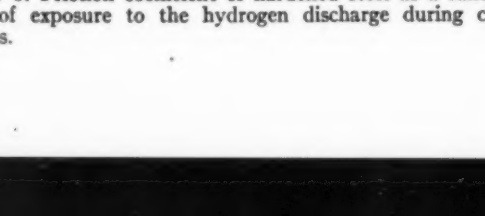


FIG. 5. Friction coefficient of hardened steel as a function of time of exposure to the hydrogen discharge during cleaning process.

average deviations of repeated determinations did not exceed about ±15 percent.

The results of measurements of hardness and shear strength for the metals used are plotted in Figs. 11 through 16.

4. Cleaning of the Samples and Effect of Surface Layers

The profound influence of surface layers of various kinds on the coefficient of friction is well known and has been amply described.[1,7] It is also known that the classical laws of friction are applicable, with friction coefficients having the customary small values of the order of unity, only when the metallic surfaces are covered with a layer of oxides or adsorbed films which prevent a clean metallic contact. If the surface of the metal is made perfectly clean (for example, through reduction of the oxides by heating in hydrogen, followed by heating and

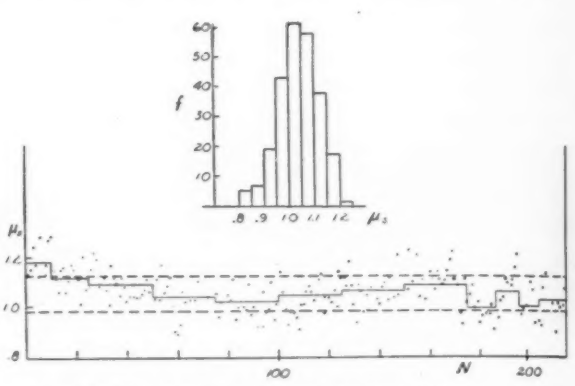


FIG. 6. Static-friction coefficient of copper (at room temperature) as a function of number of experiments performed and corresponding distribution diagram. Dotted lines represent the limits of average deviation from the mean (full line).

outgassing in high vacuum), a complete seizure results immediately when the two specimens are brought together. The unusually high "friction coefficients" (of the order of 100) observed under these conditions[8,9] have to be interpreted as a manifestation of adhesion rather than friction. Under less extreme conditions, values of the friction coefficients may assume any value within the limits already indicated.

There is, therefore, an inherent indeterminacy in the absolute value of the friction coefficient (measured as a ratio of the tangential to the normal force) in the case of very clean metals. Indeed, it is unlikely that there would be anything like an absolute value of such a coefficient at all.

In order to obtain data of some practical bearing, it was decided to perform the experiments on clean, degreased samples, partially outgassed in a high vacuum

[7] W. Claypoole and D. B. Cook, J. Franklin Inst., **233**, 453 (1942).

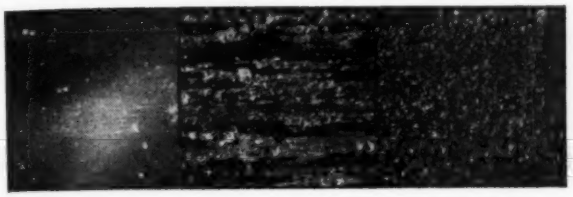
[8] Gwathmey, Leidheiser, Jr., and Smith, N.A.C.A. Tech. Note No. 1461 (1948).

[9] F. P. Bowden and J. E. Young, Nature **164**, 1089 (1949).

at temperatures which were not high enough to decompose the oxide layer. Surfaces conditioned in this manner are characterized by friction coefficients of the order of unity.

After polishing, the samples were washed with water, chloroform, and acetone, then heated in high vacuum at ~250°C for 24 hours and used for measurements immediately after cooling down. The heating, however, could not be used with the lead, tin, zinc, and hardened steel samples. Instead, bombardment with ions of an electric discharge in purified hydrogen at a pressure of about 0.1 mm mercury was used. The applied voltage was of the order of 10 kilovolts (dc) and the current was of the order of 20 milliamp; the sample itself was the cathode. When the coefficient of friction was measured as a function of the time of duration of the discharge, a gradual increase of friction was observed (Fig. 5). When cleaning the surface, the treatment was continued until the friction coefficient was considered to be sufficiently close to the limit towards which it approached asymptotically.

Some consideration was also given to the possible influence which might be produced by the condensation



A. Polished surface B. After 50 slides C. After 620 slides

FIG. 7. Change of the surface roughness of the copper sample of Fig. 6 with the number of experiments (magnification 22.5×).

of residual gases on the sample when working at the lowest temperatures. In the vacuum apparatus the sample is the coolest part of the system and, therefore, all gases and vapors will be preferentially condensed on its surface. In fact, it was observed, during an experiment at liquid-helium temperature, that the static-friction coefficient of a nickel sample dropped substantially (from 0.65 to less than 0.10) when the sample became covered with a layer of condensed gas. The amount of gas which was admitted into the apparatus in this particular case was relatively large, for the condensed layer was visible as a white film. Later experiments on nickel in sealed glass tubes filled with pure nitrogen or helium at various pressures did not show any observable change of static friction (except for what could be accounted for by the inherent temperature dependence of the metallic friction itself). The amount of gas introduced, however, was never sufficient to produce a visible condensed layer and, in fact, it may have condensed primarily on the glass walls rather than on the sample. Finally, some experiments were performed with the nickel samples immersed directly,

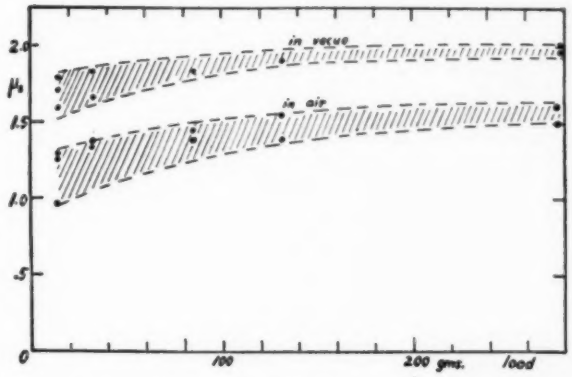


FIG. 8. Change of the static-friction coefficient of copper with weight of the slider.

either in liquid nitrogen or liquid helium. The results were as follows. (See Table II.)

Some "lubricating" effect of liquid nitrogen and liquid helium is observed, although it is relatively small. A layer of solidified gas, however, can presumably

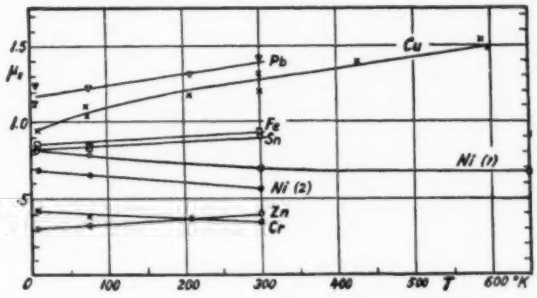


FIG. 9. Static-friction coefficients of pure metals as a function of temperature.

prevent to some extent a clean metallic contact, and the friction can thereby be reduced appreciably.

5. Polishing of the Samples and Effect of Surface Roughness

The surfaces of the samples were smoothed with a sequence of graded abrasive cloths, finished with

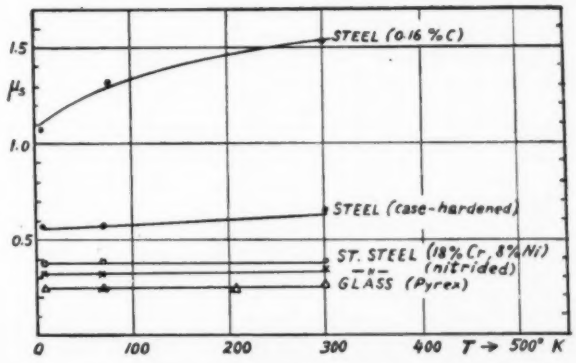


FIG. 10. Static-friction coefficients of steels and glass as a function of temperature.

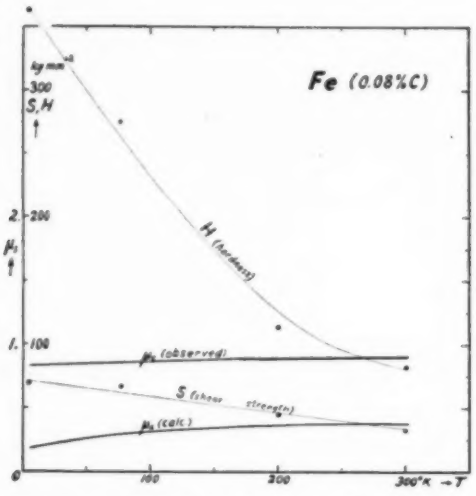


FIG. 11. Friction coefficient, hardness and shear strength of iron in the low temperature region.

"crocus cloth" and polished with wetted levigated alumina. Water was the lubricating agent in all cases. All operations were performed by hand in order to avoid excessive heating. No attempt was made to improve the smoothness by electropolishing or other means, as it was found that some wear and roughening of the surface resulted inevitably in the course of experiments. To check this procedure, an attempt was made to ascertain the influence of the surface-roughness on the friction coefficients observed. Tests were performed on a copper sample in the vacuum apparatus at room temperature.

An initial series of 220 single observations of the μ_s was taken at room temperature, starting with a freshly polished surface. The weight of the slider was 28 g. A plot of all observed points and the corresponding distribution diagram (Fig. 6) shows a fair randomness of the data. A very small decrease of friction may perhaps be noticed during the first phase of the series. The photographs in Fig. 7 show the accompanying change of the appearance of the surface (the last picture was taken after completion of a total of 650 slides). It is observed that a very marked increase of surface roughness did not cause any measurable change in the static-friction coefficient within the limits of observational errors. In spite of this rather fortunate circumstance, it was decided to take friction data only after a number of "blank" slides had been performed, in order to produce some initial roughness of the surface. The roughening of the surface was, of course, most pronounced with soft metals, such as lead or copper, and practically negligible with the hardest ones, such as case-hardened or nitrided steels.

In this connection, experiments were performed to ascertain the influence of the weight of the slider. Very light sliders should have the advantage of causing only slight wear and roughening, but on the other hand, they might not be able to penetrate sufficiently through the supposed layer of oxide on the surface. Heavy sliders should behave in just the opposite way. A typical plot of μ_s *vs* the weight of the slider is shown in Fig. 8. It turned out to be impracticable to work with weights sufficiently heavy to approach the limit of μ_s because of the high rate of wear. The weights of most of the sliders used in the work as a whole were between 20 and 50 grams.

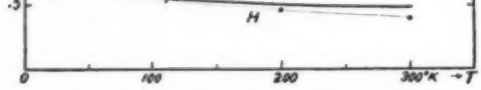


FIG. 12. Friction coefficient, hardness and shear strength of nickel in the low temperature region.

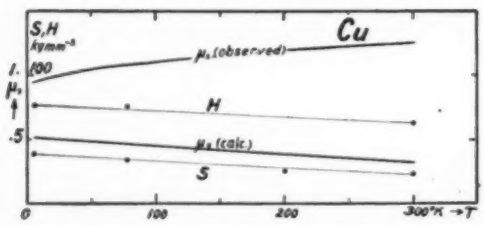


FIG. 13. Friction coefficient, hardness and shear strength of copper in the low temperature region.

RESULTS OF MEASUREMENTS AND CONCLUSIONS

The results for the static friction of pure metals are presented in Fig. 9; those of plain carbon steels and alloy steels in Fig. 10. One curve for a Pyrex-glass specimen is included in Fig. 10.

The following conclusions can be drawn from the data in Figs. 9 and 10. (a) Temperature coefficients of static friction of metals are rather small and for most metals are positive. (b) The harder the surface, the smaller is the friction coefficient and the smaller is its temperature dependence.

The highest coefficients of friction were observed with lead and copper, both of which are soft metals with high tendency to galling (see Fig. 7). These are also the metals for which the largest discrepancy between the

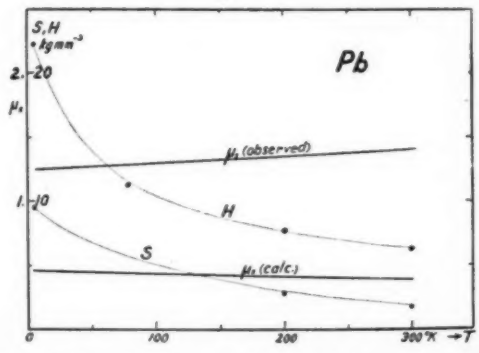


FIG. 14. Friction coefficient, hardness and shear strength of lead in the low temperature region.

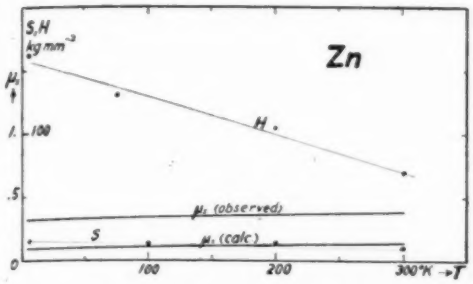


FIG. 15. Friction coefficient, hardness and shear strength of zinc in the low temperature region.

μ_s *vs* T-curve and (S/H) *vs* T-curve is found (see Figs. 13 and 14). An appreciable part of the total observed friction with these soft metals is probably due to a transfer or ploughing-type process which is neither temperature independent (as the roughness-type) nor does it obey the (S/H) *vs* T-dependence.

Lower friction coefficients are observed with metals which have either low shear strength, high hardness, or both. This is in accord with the adhesion hypothesis of friction. With hard metals the plastic deformation in contact spots is small. Therefore, the adhesion term is small and most of the remaining friction is probably due to the surface roughness. The observed friction coefficient becomes, in this case, almost temperature independent. This is seen very distinctly in the case of low carbon steel, when it has been surface hardened (Fig. 10).

A similar large decrease in friction should be observed in the case of stainless steel (when in the nitrided state as contrasted with the normal state). Here, however, the friction was small even in the original state, which may have been due to the oxide layer. The surface film (presumably, chromium oxide), which, is very stable in this case, was probably not completely removed during the cleaning process and the weight of the slider (20 g) may not have been sufficient to break through it. A similar behavior was observed with chromium. The cleaning procedure (discharge in rarefied hydrogen) was the same for all samples in Fig. 10.

With glass, which had probably the smoothest surface (fire-polished samples), the friction coefficient had the least value observed throughout the whole investigation and was completely independent of temperature.

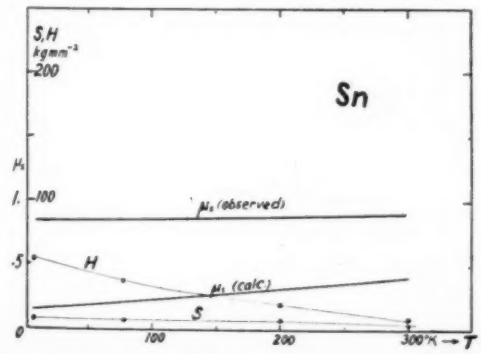


FIG. 16. Friction coefficient, hardness and shear strength of tin in the low temperature region.

Further information on the nature of the friction process was obtained by measuring the shear strength S and the hardness H as functions of temperature, and plotting the ratio S/H together with the observed friction coefficient μ_s *vs* temperature. An inspection of Figs. 11 through 16 shows that there is a large variation in the effect of temperature upon S and H for the various metals investigated. Therefore, any slope and any sign of the (S/H) *vs* T-curve are *a priori* conceivable. It turns out, however, that the *ratio* (S/H) is a rather slowly varying function of temperature and the magnitude and sign of the observed slope correspond fairly well with the slope of the friction *vs* temperature curve. The agreement is not very good in the case of tin (Fig. 16) and in the case of copper (Fig. 13) even the signs of the slopes are reversed. This discrepancy in the latter case is probably related to the transfer or ploughing type of friction observed in this case, as was pointed out above.

For the rest of the metals under investigation, the following behavior is observed in general: (a) The measured coefficients of static friction are *always higher* than the ratio S/H. (b) The slope of the measured temperature dependence of friction is approximately the same as that of the (S/H) *vs* T-curve.

These facts seem to indicate that a part of the total observed friction is due to some temperature-dependent process, whereas another part is caused by a temperature-independent component of friction, such as that attributable to surface-roughness. However, it has to be kept in mind, that neither the shear strength S nor the hardness H, as measured here, correspond exactly to what is postulated in the adhesion theory. In particular, the hardness H was measured by means of the penetration of a hard body (steel sphere) into a relatively soft metal, whereas the hardness involved in the friction of two like metals should be judged from the plastic deformation of two bodies made of the same metal. A similar objection can be raised against the way in which shear strength was measured. As a consequence, a calculated value of the ratio (S/H) cannot be expected to check with an actually observed value of the friction coefficient.

The observation under (b) substantiates the hypothesis that the temperature-dependent component of friction is due to an adhesion term of the (S/H)-type. If, of course, another temperature-dependent component of friction (e.g., the ploughing type) were present, the

TABLE II. Static friction coefficients of nickel.

T°K	μ_s in vacuum	μ_s in liquid
300	0.69	...
77	0.77	0.65 (in N_2)
4.2	0.82	0.51 (in He)

relative proportion of that component involved in the total observed friction could not be stated.

Finally, the curves of Figs. 11 through 16 contain some information concerning the plastic properties of polycrystalline metals at very low temperatures. The hardness of most of the metals investigated underwent an appreciable increase on cooling down from room temperature to the temperature of liquid helium. The hardness phenomenon, however, is too complicated to permit one to draw any detailed conclusions from the observed data. A little more promising is the temperature dependence of shear strength. It has been known, since the investigations of Boas and Schmidt,[10] that the shear strengths of single crystals of a few metals (Zn, Cd, Bi) do not increase appreciably down to the temperature of liquid helium. Such metals possess lattice structures with a single preferred slip plane, such as the basal plane in the hexagonal, close-packed structure (Zn, Cd). The results obtained on single crystals of these metals are, therefore, free of the complication caused by the mutual hindering of the individual slip planes, if more of them are equally effective. Accordingly, it was conjectured that the elementary slip process, responsible for the observed shear, is altogether temperature independent. It is apparent from our results on *polycrystalline* materials, that this is the case here also, at least for both the low crystalline-symmetry metals, zinc (hexagonal) and tin (tetragonal), Figs. 15, 16. The shear strength of both these metals is low and does not change with temperature. A rather small increase of shear strength is observed even with the polycrystalline metals of the cubic group, Cu, Ni (face-centered cubic), and Fe (body-centered cubic). The lead (face-centered cubic), Fig. 14, shows an exceptionally high increase in shear strength, resembling a fluid viscosity type (activation-energy) temperature dependence.

[10] W. Boas and E. Schmidt, Z. Physik **61**, 767 (1930).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

Optimum Source Size for the Mach-Zehnder Interferometer

F. D. BENNETT
Ballistic Research Laboratories, Aberdeen Proving Grounds, Maryland
(Received July 28, 1950)

A vector analytic treatment is given of the formation of undisturbed fringes by an ideal Mach-Zehnder interferometer with an extended source. The path difference of two interfering rays at an arbitrary point in the field is found to depend, in a simple way, upon the source point from which the rays originate, the field point examined, and a dyadic which is a function of the unit normals to the last mirror and divider plate of the interferometer. The fringe clarity condition, that all pairs of interfering rays reaching the field point have path differences within a specified range, is developed in the form of an inequality. Analysis of this fundamental inequality shows that all admissible area sources must be areas enclosed between two conics in the source plane. For fringes perpendicular to the plane of centers of the interferometer elements, the optimum source is a circle with radius inversely proportional to the square root of the number of clear fringes desired. This result holds for all interferometers of parallelogram planform. For fringes parallel to the plane of centers, each interferometer orientation presents a special case. The optimum source area is obtained for two of these, *viz.*, the 45° and 30° interferometers.

I. INTRODUCTION

IN recent years much attention has been devoted to the Mach-Zehnder interferometer as an instrument suitable for the exploration of supersonic flows. A paper by J. Winckler[1] reports the application of this instrument to the study of an axially symmetric supersonic jet of air, and in describing the theory of the interferometer, the author remarks (p. 309), "The complete theory of the formation of interference fringes from a large, extended source by the Mach interferometer has not been given so far." At about the same time, a Naval Ordnance Laboratory report by Ernest Winkler[2] appeared in which a thorough study of the Mach-Zehnder interferometer was made, employing synthetic geometrical methods and many geometrical figures and diagrams. While E. Winkler's work gives a partial answer to the question of optimum source size, his synthetic technique is complicated and hard to follow. Considerable effort must be made by the reader in visualizing the conditions described.

In this paper, a vector analytic treatment of the ideal Mach-Zehnder interferometer will be given. The problem of optimum source size and shape will be seen to be completely solvable for a large class of interferometers; furthermore, the general results will be obtained in concise, mnemonic form.

II. GENERAL VECTOR EQUATIONS

Several results from vector analytic geometry will be collected here for convenience.

In three-space, the equation of a plane with unit normal $\bar{n}$ passing through point $\bar{x}_0$ is

$$(\bar{x}-\bar{x}_0)\cdot\bar{n}=0. \tag{1}$$

[1] J. Winckler, Rev. Sci. Instr. **19**, 307 (1948).
[2] E. H. Winkler, "Analytical studies of the Mach-Zehnder interferometer," N.O.L. Report No. 1077 (December 5, 1947).

The equation for a line passing through point $\bar{x}_p$ in the direction of the unit vector $\bar{r}$ is

$$\bar{x}-\bar{x}_p=s\bar{r}, \tag{2}$$

where s is the magnitude of the distance from $\bar{x}_p$ to $\bar{x}$. The perpendicular distance from the point $\bar{x}_p$ to the plane (1) is found by replacing $\bar{r}$ in (2) by the normal $\bar{n}$ and requiring that $\bar{x}$ be the point of intersection of the line with the plane; thus

$$s_\perp=(\bar{x}_0-\bar{x}_p)\cdot\bar{n}. \tag{3}$$

To treat the reflection of light from a plane mirror, we need to obtain the unit vector of the outgoing ray $\bar{r}_j$ in terms of the unit vector of the incoming ray $\bar{r}_{j-1}$ and the normal to the plane $\bar{n}_j$. The laws of reflection are expressed by the equations

$$(\bar{r}_j\cdot\bar{n}_j)=-(\bar{r}_{j-1}\cdot\bar{n}_j),$$

and

$$\bar{r}_j=a\bar{r}_{j-1}+b\bar{n}_j.$$

These with the requirement that $\bar{r}_j^2=1$ yield

$$\bar{r}_j=\bar{r}_{j-1}-2(\bar{r}_{j-1}\cdot\bar{n}_j)\bar{n}_j. \tag{4}$$

By exactly similar reasoning, one could express the normal $\bar{n}_j$ in terms of $\bar{r}_j$ and $\bar{r}_{j-1}$, assuming these to be known.

III. SPECIFICATION OF THE INTERFEROMETER

We shall consider only interferometers made of ideal elements. The collimating and camera lenses are assumed to be perfect thin lenses. The dividing plates and mirrors are assumed to be perfectly plane and of zero thickness, producing no changes other than beam division and/or change of direction of the incident light.

To specify a standard interferometer configuration, consider a plane rectangle of sides a, b in the x, y plane with its lower left-hand corner at the origin and sides parallel to the x and y directions. The elements of the interferometer are located, as in Fig. 1, with their centers on the corners of the rectangle, and lie in vertical planes whose traces have slope -1. Thus, we have a 45° interferometer in standard position with element ①, the first divider plate, at (o, a); element ②, a reflecting mirror, at (b, a); element ③, reflecting mirror, at (o, o); and element ④, the second divider plate, at (b, o). The unit normals $\bar{n}_1$ and $\bar{n}_3$ corresponding to elements ① and ③ lie in the x, y plane at 45° with the x axis and pointing into the first quadrant, while the unit normals $\bar{n}_2$, $\bar{n}_4$ are antiparallel with $\bar{n}_1$, $\bar{n}_3$. Parallel light entering the interferometer from the collimating lens at $(o, a+l)$ proceeds in the negative y direction to ① where it is divided into two beams which pass through the interferometer along opposite sides of the rectangle, and which recombine at ④, there to exit again in the negative direction.

If the source illuminating the interferometer consists of a single point on the y axis at $(o, a+l+f)$, where f is the focal length of the lens, the collimator will form a single parallel beam which may be characterized by the unit vector $\bar{r}_c$ located at the center of the lens and pointing in the negative y direction. This type of source is that which is commonly employed in discussing the elementary theory of the interferometer. The $\bar{r}_c$ unit vector has the property of lying at the center of the beam and pointing along its axis; thus, it is capable of representing each ray in the beam. Furthermore, planes with $\bar{r}_c$ as normal intersect the beam in surfaces of constant phase.

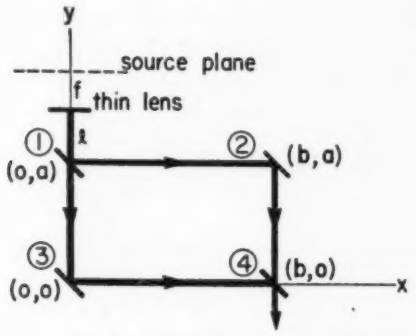


FIG. 1. Standard interferometer configuration.

If an extended area source in the focal plane of the lens be considered, it is clear that the lens produces a parallel beam corresponding to each point on the source, and characterized by the unit vector $\bar{r}$, with origin at the center of the lens, directed along the line joining the source point with the lens center. A source of arbitrary shape is seen to give rise to a bundle of unit vectors whose tips lie on the unit sphere with origin at the center of the lens. In considering the formation of fringes by the interferometer, it is helpful to regard this bundle of unit vectors as if it were a luminous object replacing the lens and source altogether.

The effect of the interferometer is to split each beam arising from a single source-point into two coherent beams of parallel light, allowing them to traverse equal paths before recombination at plate ④.

In terms of the characteristic unit vectors $\bar{r}$, the interferometer produces two virtual bundles of unit vectors at a distance $L=a+b+l$ above plate ④. With ④ at 45°, these virtual images exactly overlap and no fringes are formed. If ④ is rotated through a small angle $\epsilon/2$ leaving its center fixed, one of the virtual images is rotated about the same center on a radius L, through the angle ϵ. In general, the bundle of vectors is also rotated like a rigid body about line L as the axis of rotation.

Since the reflection from ② fixes the position of the unrotated bundle of unit vectors, the representative vector denoted by $\bar{r}$ at the lens will become $\bar{r}_2$ at mirror ② in conformity with the notation of Eq. (4). As the reflection from ④ fixes the position of the rotated bundle, its representative vector will become $\bar{r}_4$. The result then is two diverging pencils of parallel beams so oriented that the central beams of each, characterized by virtual $\bar{r}_{2c}$ and its displaced image $\bar{r}_{4c}$, converge toward the center of plate ④.

While in practice, fringes are rarely obtained by adjustment of plate ④ alone, for the purposes of the analysis it is most convenient to regard all other elements as fixed, and rotate ④. As will be seen later, the analysis of this case can be applied without modification to a large class of adjustments involving three elements of the interferometer.

IV. DERIVATION OF THE PATH DIFFERENCE d

The central problem before us is to examine at an arbitrary field point $\bar{x}_p$ (with origin now determined at the center of plate ④), the phase of any pair of interfering rays reaching the point. This we may do by recalling that the first divider plate produces two and only two beams from each point on the source. Any pair of rays from these two beams may combine to cause interference. Rays selected from beams belonging to different source points will not interfere because of lack of coherence. Since we have already seen that the unit vector $\hat{r}$ at the lens splits into $\hat{r}_2$ and $\hat{r}_4$, we have in these vectors the identification of corresponding pairs of coherent beams. Rays proceed parallel to these unit vectors in the respective beams; we have only to determine the perpendicular distance from point $\bar{x}_p$ to the initial planes of equal phase, *viz.*, the planes passing through the origins of the virtual bundles of vectors. The difference of these distances is the path difference d, which governs the phase of the interference of the two rays.

To find these origins, we utilize the fact that the virtual bundles of vectors appear at distance $-L$ from the center of ④ in the $\hat{r}_{2c}$ and $\hat{r}_{4c}$ directions, thus

$$\bar{x}_2 = -L\hat{r}_{2c},$$

and

$$\bar{x}_4 = -L\hat{r}_{4c}, \tag{5}$$

from Eq. (2).

The perpendicular distance from $\bar{x}_p$ to the plane passing through $\bar{x}_2$ with normal $\hat{r}_2$ is by (3)

$$s_2 = (\bar{x}_2 - \bar{x}_p)\cdot\hat{r}_2;$$

similarly for the plane passing through $\bar{x}_4$ with normal $\hat{r}_4$

$$s_4 = (\bar{x}_4 - \bar{x}_p)\cdot\hat{r}_4.$$

The path difference

$$\begin{aligned} d = s_4 - s_2 &= (\bar{x}_4 - \bar{x}_p)\cdot\hat{r}_4 - (\bar{x}_2 - \bar{x}_p)\cdot\hat{r}_2 \\ &= L(\hat{r}_{2c}\cdot\hat{r}_2 - \hat{r}_{4c}\cdot\hat{r}_4) + (\hat{r}_2 - \hat{r}_4)\cdot\bar{x}_p, \end{aligned} \tag{6}$$

using Eq. (5).

Now from (4), we find by substitution:

$$\hat{r}_j\cdot\hat{r}_{jc} = \hat{r}_{j-1}\cdot\hat{r}_{j-1c}, \tag{7}$$

which expresses the invariance of angle between two rays under plane reflection. Using (7) to trace back through the elements of the interferometer, we establish that

$$\hat{r}_2\cdot\hat{r}_{2c} = \hat{r}_4\cdot\hat{r}_{4c} = \hat{r}\cdot\hat{r}_c; \tag{8}$$

thus (6) reduces to

$$d = (\hat{r}_2 - \hat{r}_4)\cdot\bar{x}_p. \tag{9}$$

We see that the path difference expression is independent of any lengths characteristic of the interferometer so long as the origins of the virtual unit vectors lie the same distance, $-L$, behind the center of ④ in the $\hat{r}_{2c}$ and $\hat{r}_{4c}$ directions.

An immediate broadening in two directions of the class of interferometers covered by this analysis is now possible. In the first place, the rectangle may be sheared in the y direction into a parallelogram with vertical sides. The central rays formed by splitting $\hat{r}_c$ at ① traverse the sides of the parallelogram, providing that the normals of the interferometer elements are adjusted to bisect the acute angles into which they point, retaining $\bar{n}_1$, $\bar{n}_3$ parallel as before and antiparallel to $\bar{n}_2$, $\bar{n}_4$ prior to rotation of $\bar{n}_4$ to produce fringes. Such a sheared interferometer will be denoted by the angle between $\bar{n}_1$ and the vertical, e.g., a 30° interferometer is one in which the traces of the plates in standard position are inclined at 30° to the negative x direction; $\bar{n}_1$, $\bar{n}_3$ thus make 30° with the vertical.

Secondly, the origin may be moved from the center of ④ to any line y=const between ② and ④ (or exterior to this segment for that matter) still preserving the essential conditions which lead to Eq. (9). One way of doing this is to rotate $\bar{n}_1$ toward the vertical, and translate ② vertically so that the central ray from ① is still incident on the center of ② but rotating its normal so that $\hat{r}_{2c}$ is no longer vertical. Rotation of ④ in the appropriate direction will then give an intersection of $\hat{r}_{2c}$ and $\hat{r}_{4c}$, which defines the new origin.

It is easily shown that given θ and ϕ, the angles through which $\bar{n}_2$ and $\bar{n}_4$ are to be rotated, and y_0, the vertical distance to the new origin measured from the center of ④, there always exists a real solution to the problem of specifying α the angle of rotation of $\bar{n}_1$ (and the vertical translation of ② which depends upon α) in such a way that Eqs. (5) are preserved with an appropriate value of the parameter L.

V. ANALYSIS OF THE PATH DIFFERENCE FORMULA

To develop Eq. (9) further, let us trace the representative (beam) vector $\hat{r}$ through the interferometer using (4). Since $\bar{n}_1 = -\bar{n}_2$, we have

$$\hat{r}_2 = \hat{r}, \tag{10}$$

$$\hat{r}_3 = \hat{r} - 2(\hat{r}\cdot\bar{n}_3)\bar{n}_3, \tag{11}$$

and

$$\hat{r}_4 = \hat{r}_3 - 2(\hat{r}_3\cdot\bar{n}_4)\bar{n}_4. \tag{12}$$

Substituting these equations into (9) gives

$$d = 2\hat{r}\cdot D\cdot\bar{x}_p, \tag{13}$$

where

$$D = [\bar{n}_3\bar{n}_3 - 2\bar{n}_3(\bar{n}_3\cdot\bar{n}_4)\bar{n}_4 + \bar{n}_4\bar{n}_4]. \tag{14}$$

D is a dyad or matrix and in (13) is operated on from the left by dotting in $\hat{r}$, from the right by dotting in $\bar{x}_p$

This remarkably simple result allows us to express the path difference of interfering rays from source point $\bar{r}$ at any field point $\bar{x}_p$ providing we know the unit normals $\bar{n}_3$, $\bar{n}_4$ of the interferometer.

It is immediately clear from the form of D that there is no contribution to d from any component of $\bar{x}_p$ in the direction $\bar{n}_4 \times \bar{n}_3$, i.e., d is constant along lines in the $\bar{n}_4 \times \bar{n}_3$ direction. This direction is therefore the direction of clear fringes formed by the interferometer. Vertical fringes* are formed when $\bar{n}_3$, $\bar{n}_4$ remain in the plane of centers, while horizontal fringes occur when $\bar{n}_4$ is rotated in a plane containing $\bar{n}_3$, and perpendicular to the plane of centers.

The character of the fringes may be seen from the following. Fringes are curves of intersection between the plane of observation (photographic plate) and surfaces of constant path difference d. For $d=$ const and a given $\bar{r}$ (source point) (13) is the equation of the family of parallel planes with normal $\bar{r} \cdot D$. As $\bar{r} \cdot D$ is a linear function of $\bar{n}_3$ and $\bar{n}_4$, the planes themselves are perpendicular to the plane of $\bar{n}_3$, $\bar{n}_4$ and thus parallel to the $\bar{n}_4 \times \bar{n}_3$ direction regardless of $\bar{r}$. For point source $\bar{r}=\bar{r}_c$, the planes of constant path difference are very closely parallel to $\bar{r}_c$, and the fringes will be parallel straight lines (not necessarily parallel to $\bar{n}_4 \times \bar{n}_3$). For a source point off the axis, the corresponding family of planes does not intersect the plate along the same straight lines as before. The separation between the fringes for $\bar{r}_c$ and $\bar{r}$ is zero at the origin and increases both linearly with x and as a function of the angle between $\bar{r} \cdot D$ and $\bar{r}_c \cdot D$; hence, with an extended source, the composite fringes, while unchanging in the $\bar{n}_4 \times \bar{n}_3$ direction, become progressively fuzzier with distance from the origin.

If the plane of observation contains $\bar{n}_4 \times \bar{n}_3$, the fringes are of course parallel to this direction. Rotation of the plane about $\bar{n}_4 \times \bar{n}_3$ causes the fringe spacing to pass through a relative minimum, as will be seen later from the fringe width Eq. (22). It is reasonable to suppose that the sharpest fringes will occur for the narrowest fringe width, i.e., at the relative minimum; since otherwise the fuzziness caused by an extended source is projected onto a wider area and contrast of the fringes is reduced.

D may be further modified to advantage if we set

$$\bar{n}_4 = a\bar{n}_3 + b\bar{\imath}, \tag{15}$$

where $\bar{\imath}$ is a unit vector in the $\bar{n}_3 \times (\bar{n}_4 \times \bar{n}_3)$ direction. (Note that $\bar{n}_3$, $\bar{n}_4 \times \bar{n}_3$, $\bar{\imath}$ form a right-handed triad.)

Since $\bar{n}_3 \cdot \bar{n}_4 = -\cos\epsilon/2$, and $\bar{n}_4^2 = 1$, we find

$$\bar{n}_4 = -\cos\epsilon/2\ \bar{n}_3 + \sin\epsilon/2\ \bar{\imath}, \tag{16}$$

and

$$D = \sin\epsilon/2[\sin\epsilon/2\ \bar{n}_3\bar{n}_3 + \cos\epsilon/2(\bar{n}_3\bar{\imath} - \bar{\imath}\bar{n}_3) + \sin\epsilon/2\ \bar{\imath}\bar{\imath}], \tag{17}$$

* Henceforth the adjective "vertical" designates fringes perpendicular to the plane of centers of the interferometer elements, while "horizontal" designates fringes parallel to this plane.

where $\epsilon/2$ is the angle of rotation of $\bar{n}_4$ to produce fringes.

Since no component of $\bar{x}_p$ in the $\bar{n}_4 \times \bar{n}_3$ direction has any effect on d, there is no loss in assuming henceforth that $\bar{x}_p$ lies entirely in the plane of $\bar{n}_3$, $\bar{\imath}$ perpendicular to $\bar{n}_4 \times \bar{n}_3$. Let

$$\bar{x}_p = \bar{x} = x(\bar{n}_3 \sin\psi + \bar{\imath} \cos\psi). \tag{18}$$

Substituting (18) and (17) into (13), we obtain

$$d = 2x \sin\epsilon/2\, \bar{r} \cdot \bar{p}, \tag{19}$$

where

$$\bar{p} = \bar{n}_3 \cos(\psi - \epsilon/2) - \bar{\imath} \sin(\psi - \epsilon/2). \tag{20}$$

$\bar{x}$ is the vector to the field point under examination in the plane normal to $\bar{n}_4 \times \bar{n}_3$. $\bar{p}$ is a unit vector in the same plane but approximately perpendicular to $\bar{x}$, as may be seen from

$$\bar{p} \cdot \bar{x} = x \sin\epsilon/2 \doteq 0. \tag{21}$$

Now $\bar{r} \cdot \bar{p}$ is the cosine of the angle ϕ between $\bar{r}$ and $\bar{p}$ and depends upon the source point represented by $\bar{r}$ and upon the vector $\bar{p}$, which is very closely normal to the direction of investigation in the field of fringes, i.e., normal to the photographic plate. Using (19), if

$$|d_2 - d_1| = \lambda = 2 \sin\epsilon/2\, |(\bar{r} \cdot \bar{p})(x_2 - x_1)|,$$

then the fringe width is

$$w = |x_2 - x_1| = \frac{\lambda}{2 \sin\epsilon/2\, |(\bar{r} \cdot \bar{p})|}. \tag{22}$$

Some further information regarding the fringe pattern may now be obtained. w takes on its smallest value for a given $\epsilon/2$ when $|\bar{r} \cdot \bar{p}| = 1$, i.e., $\bar{r}$ and $\bar{p}$ are parallel or antiparallel. As $\bar{p}$ lies in the plane of $\bar{n}_3$ and $\bar{n}_4$, and $\bar{r}$ will usually lie close to $\bar{r}_c$, the minimum w will be attained only when $\bar{r}_c$ lies in or near the $\bar{n}_3$, $\bar{n}_4$ plane, i.e., perpendicular to $\bar{n}_4 \times \bar{n}_3$. As $\bar{n}_4 \times \bar{n}_3$ lies always in the plane normal to $\bar{n}_3$, it will be perpendicular to $\bar{r}_c$ only when also perpendicular to the plane of centers. This is the condition for vertical fringes. Other fringe orientations produce relative minima larger than that for vertical fringes, e.g., for $\bar{n}_4 \times \bar{n}_3$ horizontal, $|\bar{r}_c \cdot \bar{p}| = \cos\theta$, where θ is the angle of the interferometer.

For vertical fringes with point source $\bar{r} = \bar{r}_c$, we may have $|\bar{r}_c \cdot \bar{p}| = 1$, and (22) assumes the familiar standard form

$$w_s = \lambda/2 \sin\epsilon/2 \doteq \lambda/\epsilon. \tag{23}$$

The conjecture made earlier in the paragraph, that laterally sharpest fringes are associated with minimum fringe width, leads to the conclusion that vertical fringes with $\bar{p} \cdot \bar{r}_c = 1$ will be sharpest of all. As $\bar{p}$ is normal to the $\bar{x}$ direction in the photographic plate within $\epsilon/2$, we see that the usual arrangement, where vertical fringes are observed on a plate normal to $\bar{r}_c$, is actually very close to the best possible, and (23) applies.

VI. FUNDAMENTAL INEQUALITY

Equation (22) shows the dependence of fringe width on the position of the source point represented by $\bar{r}$; thus an extended area source will produce a composite fringe system consisting of an infinite set of overlapping fringe systems. All those source points producing patterns with the same fringe width are characterized by the requirement

$$\bar{r}\cdot\bar{p}=\cos\phi=\text{const.} \tag{24}$$

(24) restricts the vectors $\bar{r}$ to lie on the cone with vertex at the center of the unit sphere and generator making angle ϕ with the unit vector $\bar{p}$. The intersection of this cone with the plane of the source, whose normal is $\bar{r}_c$, gives the source curve which produces the given fringe width. This curve is, of course, a conic section.

In order to obtain a clear fringe system, we must either require that the source consist only of points for which $d=\text{const}$ and (24) follows, or we must require that d differ by a small preassigned quantity from the predetermined constant value. The second alternative is less restrictive than the first and will lead to area-type sources which are practically realizable. Selecting d_c as the comparison value, i.e., in effect choosing the path difference for the central source point as the standard, we write

$$|d-d_c|\leqslant K\lambda, \tag{25}$$

where K is a suitably small number depending upon the clarity that we require in the fringe system.

We shall call this expression the *fundamental inequality of the Mach-Zehnder interferometer*. It is a generalization of the form used by E. H. Winkler in derivation of his Eq. (2). From it can be deduced the size and shape of an admissible source for any interferometer of the class specified in Sec. IV with any desired orientation of fringes.

While d_c has been selected as the comparison value in (25), situations may arise where some advantage may be obtained by excluding the central source point. In such a case, a suitable reference value of d should be chosen and substituted for d_c.

Returning to (19), we see that for $x=0$, $d=0$, and we have the central fringe condition for which all source points have zero path differences. Proceeding in $+$ or $-x$ direction, the path difference increases linearly as one passes through fringes 0, ±1, ±2, ±3 $\cdots$. A simplification in (19) occurs if we agree to measure x in units of standard fringe width w_s. Using (23), the expression for w_s, (19) becomes

$$d=N\lambda\bar{r}\cdot\bar{p}, \tag{26}$$

and (25)

$$|\bar{r}\cdot\bar{p}-\bar{r}_c\cdot\bar{p}|\leqslant(K/N), \tag{27}$$

where we now understand N to be positive. For most practical situations, the actual fringe width differs so little from the standard w_s that the number N gives accurately the number of fringes on either side of center for which the fundamental inequality is satisfied. In using (27), we may specify the number of fringes we desire within a given clarity criterion rather than specifying the size of the field to be investigated.

Inequality (27) is unchanged if we replace $\bar{r}$ and $\bar{r}_c$ by their negatives. The advantages are twofold: 1) $0\leqslant\bar{r}_c\cdot\bar{p}\leqslant1$, i.e., ϕ_c is an acute angle and representation of $\bar{p}$ and $\bar{r}_c$ on the unit sphere is easier to visualize; 2) $\bar{r}_c$ is now directed along the axis toward the source plane, whose equation relative to the center of the sphere is

$$(\bar{x}-f\bar{r}_c)\cdot\bar{r}_c=0. \tag{28}$$

The fundamental inequality may now be written as

$$\bar{r}_c\cdot\bar{p}-(K/N)\leqslant\bar{r}\cdot\bar{p}\leqslant\bar{r}_c\cdot\bar{p}+(K/N), \tag{29}$$

or

$$\cos\phi_c-(K/N)\leqslant\cos\phi\leqslant\cos\phi_c+(K/N). \tag{30}$$

We see that the permissible $\bar{r}$ vectors lie on the unit sphere between the two cones whose generators make angles.

$$\phi=\cos^{-1}[\cos\phi_c\pm(K/N)], \tag{31}$$

with vector $\bar{p}$. In the plane (28), the permissible source may occupy the area contained between the conics formed by the intersections of the cones with the plane.

The equation of the conic in the source plane is easily found if we employ the orthogonal triad of unit vectors

$$\bar{r}_c$$
$$\bar{\rho}=\frac{\bar{p}\times\bar{r}_c}{\sin\phi_c},$$
$$\bar{\eta}=\frac{\bar{r}_c\times(\bar{p}\times\bar{r}_c)}{\sin\phi_c}, \tag{32}$$

and construct the vector

$$\bar{R}=f\bar{r}_c+x\bar{\eta}+y\bar{\rho}, \tag{33}$$

which satisfies the equation of the plane (28). The requirement that $\bar{R}$ be a generator of the cone is

$$(\bar{R}\cdot\bar{p})/|\bar{R}|=\cos\phi; \tag{34}$$

whence using (32), (33), and simplifying, we have the standard form

$$\bar{X}^2/a^2+\bar{Y}^2/b^2=1, \tag{35}$$

with

$$\left.\begin{aligned}\bar{X}&=x-\frac{f\sin\phi_c\cos\phi_c}{\cos^2\phi-\sin^2\phi_c}\\ a^2&=\frac{f^2\sin^2\phi\cos^2\phi}{(\cos^2\phi-\sin^2\phi_c)^2}\\ b^2&=\frac{f^2\sin^2\phi}{\cos^2\phi-\sin^2\phi_c}\\ \frac{b^2}{a^2}&=\frac{|\cos^2\phi-\sin^2\phi_c|}{\cos^2\phi}\end{aligned}\right\}. \tag{36}$$

The conic is an ellipse, parabola, or hyperbola as the discriminant

$$\Delta=(\cos^2\phi-\sin^2\phi_c)\gtreqless 0. \tag{37}$$

As K/N is usually small, ϕ will be close to ϕ_c, the $\Delta=0$ value will appear when $\phi_c\doteq\pi/4$. This situation arises, as will be seen later, when one considers the formation of horizontal fringes in the 45° interferometer.

VII. VERTICAL FRINGES

For vertical fringes $\bar{n}_4\times\bar{n}_3$, the fringe direction, is perpendicular to the plane of centers of the interferometer. $\bar{n}_3$, $\bar{\imath}$, and consequently $\bar{p}$, all lie in the plane of centers. Since $\bar{p}$ is nearly perpendicular to the $\bar{x}$ direction of investigation, $\bar{p}$ will be almost parallel with $\bar{r}_c$ when $\bar{x}\cdot\bar{r}_c=0$, i.e., when the normal to the photographic plate is $\bar{r}_c$. More precisely, $\bar{r}_c\cdot\bar{p}=1$ implies

$$(\bar{r}_c\cdot\bar{n}_3)\cos(\psi-\epsilon/2)-(\bar{r}_c\cdot\bar{\imath})\sin(\psi-\epsilon/2)=1,$$

or

$$\cos(\psi-\epsilon/2-\theta)=1, \tag{38}$$

where θ is the angle of the interferometer. Equation (38) is satisfied if $\psi=\theta+\epsilon/2$; in other words, if the photographic plate is rotated through a positive angle $\theta+\epsilon/2$ from the $\bar{\imath}$ direction which lies along the trace of the mirror ③. The plate is thus seen to be parallel to the bisector of the angle between the $\bar{r}_{2c}$ and $-\bar{r}_{4c}$ vectors. This is the condition which E. H. Winkler has given for the plane of sharpest fringes. In this derivation it appears as the plane for which the largest source may be used to obtain fringes of a given clarity; for with $\bar{r}_c\cdot\bar{p}=1$ the fundamental inequality becomes

$$1-(K/N)\leqslant\bar{r}\cdot\bar{p}\leqslant 1\leqslant 1+(K/N); \tag{39}$$

hence the admissible source is contained within the circle whose radius is

$$\rho=f\tan\phi_L, \tag{40}$$

where $\cos\phi_L=1-(K/N)$. Neglecting K^2/N^2 and higher powers, this relation becomes

$$\rho=f(2K/N)^{\frac{1}{2}}, \tag{41}$$

which is equivalent to Winkler's Eq. (2). It should be remarked that the argument given in deriving (41) is entirely independent of the angle of the interferometer and thus applies to any in our general class. That (41) gives the largest permissible source is seen from the following heuristic argument: as $\bar{r}_c\cdot\bar{p}$ departs from value 1, the central cone is no longer degenerate. For small deviations, the excluded area near the center increases approximately as $f^2\xi^2$, where ξ is the angle between $\bar{r}_c$ and $\bar{p}$. In the meantime, the outer area has remained practically constant; since, although the circle has become an ellipse of area πab, a, and b are equal to within infinitesimals of the second order. The total available source area is seen to decrease. Aperture considerations restrict the usable source to the vicinity of the origin and so exclude the extreme case with $\xi\doteq\pi/4$, where the inner curve in the source plane is an ellipse while the outer curve is one branch of an hyperbola. Here the "interior" area is clearly infinite, but the usable region near the origin is extremely narrow and much less in area than that of the circle given above.

That (41) imposes a very weak restriction on source size may be seen by setting $K=\frac{1}{10}$ and $N=200$ fringes. For a 2-in. lens with $f=8$ in., the diameter of the source is somewhat larger than half an inch. Practical spark sources used in high speed interferometry have not approached this size, to the author's knowledge.

VIII. HORIZONTAL FRINGES

It is possible to discuss the formation of fringes of any orientation by means of the analysis given above; however, we shall mention here only horizontal fringes. This case, which is representative of arbitrary orientations, is much more difficult to treat because the results depend strongly on the angle of the interferometer, e.g., whether it is a 45° or 30° interferometer has a strong effect on the permissible source.

As before, $\bar{n}_4\times\bar{n}_3$ defines the fringe direction, and, as this is taken to be horizontal, it is clear that $\bar{n}_3$ and $\bar{n}_4$ and $\bar{\imath}$ all lie in the vertical plane. $\bar{n}_4$ may be rotated vertically out of the plane of centers through the angle $\epsilon/2$.

In order to obtain lengthwise clear fringes, it is necessary to orient the photographic plate to include the $\bar{n}_4\times\bar{n}_3$ direction, i.e., it must be made to include the trace of mirror ③. Since $\bar{r}_c$ makes angle θ with $\bar{n}_3$ in the horizontal plane of centers, and $\bar{p}$ must lie in a vertical plane containing $\bar{n}_3$, the smallest angle between $\bar{r}_c$ and $\bar{p}$ (consequently the largest source) is obtained when $\bar{n}_3\cdot\bar{p}=1$. Using (20), the expression for $\bar{p}$, as before, we obtain

$$\cos(\psi-\epsilon/2)=1, \tag{42}$$

whence $\psi=\epsilon/2$ and the condition for clearest fringes (as given by E. H. Winkler) is to rotate the photographic plate around $\bar{n}_4\times\bar{n}_3$ through $\epsilon/2$ from vertical. In this case, the plate is normal to $\bar{n}_4$.

Even in this favorable case, $\cos\phi_c=\bar{r}_c\cdot\bar{p}=\cos\theta$, and unless the interferometer chosen is one of extreme shear for which $\theta\doteq 0$, the angle between $\bar{r}_c$ and $\bar{p}$ is appreciable.

In order to obtain an idea how the permissible source for a given interferometer appears, it is convenient to obtain approximate expressions for the constants of the conics neglecting terms in K^2/N^2 and higher. While the following equations are general, for horizontal fringes $\phi_c=\theta$ will hold.

Since the limiting values of $\cos\phi$ are given by the fundamental inequality, we investigate only those curves for which

$$\cos^2\phi=(\cos\phi_c\pm K/N)^2, \tag{43}$$

where in this and the following expressions the upper sign refers to the upper limit, etc. The discriminant Δ

in (37) becomes

$$\Delta \doteq \cos 2\phi_c \left(1 \pm g \frac{\cos\phi_c}{\cos 2\phi_c}\right) \gtrless 0, \tag{44}$$

with $g = 2K/N$.

We find for the center of the conic (35)

$$x \doteq \frac{f/2 \ \tan 2\phi_c}{1 \pm (g \cos\phi_c / \cos 2\phi_c)}, \tag{45}$$

$$\doteq f/2 \ \tan 2\phi_c \left(1 \mp \frac{g \cos\phi_c}{\cos 2\phi_c}\right), \tag{46}$$

when $(g \cos\phi_c/\cos 2\phi_c) \ll 1$. Similarly,

$$a \doteq \pm f/2 \ \tan 2\phi_c \frac{\left(1 \mp \dfrac{g \cot 2\phi_c}{\sin\phi_c}\right)}{\left(1 \pm \dfrac{g \cos\phi_c}{\cos 2\phi_c}\right)}, \tag{47}$$

$$\doteq \pm f/2 \ \tan 2\phi_c \left[1 \mp \frac{g(1+\cos^2 2\phi_c)}{\sin\phi_c \sin 4\phi_c}\right], \tag{48}$$

$$b \doteq \pm \frac{f}{(2)^{\frac{1}{2}}} \left(\frac{1-\cos 2\phi_c}{\cos 2\phi_c}\right)^{\frac{1}{2}} \frac{[1 \mp (g \cos\phi_c / 2 \sin^2\phi_c)]}{[1 \pm (g \cos\phi_c / 2 \cos 2\phi_c)]}, \tag{49}$$

$$\doteq \pm \frac{f}{(2)^{\frac{1}{2}}} \left(\frac{1-\cos 2\phi_c}{\cos 2\phi_c}\right)^{\frac{1}{2}} \left[1 \mp \frac{g \cos^3\phi_c}{2 \sin^2\phi_c \cos 2\phi_c}\right], \tag{50}$$

and

$$\frac{b}{a} \doteq \frac{(|\cos 2\phi_c|)^{\frac{1}{2}}}{\cos\phi_c} \frac{[1 \pm (g \cos\phi_c/\cos 2\phi_c)]^{\frac{1}{2}}}{[1 \pm (g \cos\phi_c/\cos^2\phi_c)]^{\frac{1}{2}}}, \tag{51}$$

$$\frac{b}{a} \doteq \left(\frac{|\cos 2\varphi_c|}{\cos\varphi_c}\right)^{\frac{1}{2}} \left[1 \pm \frac{g(1-\cos 2\phi_c)}{4 \cos\phi_c \cos 2\phi_c}\right]. \tag{52}$$

Using these equations and assuming $\theta = \pi/4$, a 45° interferometer, we find that the permissible source of $K = \frac{1}{10}$, $N = 200$ is the area contained between an ellipse and an hyperbola in the source plane. The constants of these figures are as follows:

Ellipse: $x \doteq 707f$

$a \doteq \pm 707f$

$b/a \doteq \dfrac{1}{(707)^{\frac{1}{2}}} \doteq \dfrac{1}{27}$

Hyperbola: $x \doteq -707f$

$a \doteq \pm 707f$

with asymptotes $y \doteq \pm 1/(707)^{\frac{1}{2}} x$.

The result is the extremely narrow, nearly horizontal, ribbon area imprisoned in the first and fourth quadrants between the conics given above. The origin is, of course, included in the source.

For the 30° interferometer, $\theta = \pi/6$, both conics are ellipses with the following constants:

$$x \doteq (3)^{\frac{1}{2}}/2f(1 \mp (3)^{\frac{1}{2}} g),$$
$$a \doteq (3)^{\frac{1}{2}}/2f(1 + 5(3)^{\frac{1}{2}}/3 \ g),$$
$$b/a \doteq (2/3)^{\frac{1}{2}}(1 \pm (3)^{\frac{1}{2}}/6 \ g),$$

from which it is clear that with $g = 10^{-3}$ as before, the ellipses are not far from circular as $b/a \doteq 0.816$. The width of the source on the axis is about $0.014f$. This compares with the diameter $2(g)^{\frac{1}{2}}f = 0.063f$ for the case of vertical fringes. Aperture limitations set by plate ④ or the camera lens are likely to restrict the useful source to less than 0.1 radian in angular width, in which case the circular source of 0.063 radian angular diameter will have a clear advantage in useful area.

The practical situation of interest with horizontal fringes is, of course, that for which the photographic plate is perpendicular to $\bar{r}_c$. Since this plane no longer contains the $\bar{n}_4 \times \bar{n}_3$ direction, there will be a non-zero component of $\bar{x}_p$ in the plane normal to $\bar{n}_4 \times \bar{n}_3$ for all points except the origin. Investigating the boundaries of a circular field in the photographic plate is equivalent to exploring an ellipse in the plane perpendicular to $\bar{n}_4 \times \bar{n}_3$. The $\bar{p}$ vector will be nearly normal to the arbitrary vector on this ellipse.

For each $\bar{p}$ vector (i.e., for each $\bar{x}$ direction in the plane of the ellipse) the permissible source will lie between the two conics defined by the upper and lower bounds of the fundamental inequality. As there will be an infinity of different $\bar{p}$ vectors, there will be a corresponding infinity of pairs of conics. The permissible source is thus the common area between all pairs of conics in the infinite set. To determine this area is a complicated calculation which we shall not attempt here. Suffice it to say that the resulting source will be much more restricted than in any of the previous examples.

IX. ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. J. H. Giese and Mr. D. B. Sleator for many helpful discussions of this work.

Periodic Statistical Distortion of Unidirectionally Ordered Diffractors, with Application to Collagen*

RICHARD S. BEAR AND ORVIL E. A. BOLDUAN†
Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received June 16, 1950)

The diffraction expected of distorted, one-dimensionally ordered cylinders (fibrils) is considered quantitatively. The distortion involved is statistically of cylindrical symmetry about any point in the fibril, and periodic along the fibril axis. An important case is that of the "mixed perfect and imperfect fibril," in which, at some axial levels, distortion is absent, and at others it is appreciable. For such systems it is shown that the reciprocal-space disk corresponding to a given diffraction layer line may be regarded as composed of three sub-disks: one of perfection, whose central intensity is maximal and whose diameter is small and independent of layer-line index; another of longitudinal or axial imperfection, whose central intensity is also maximal; and a third, related to radial imperfection, whose intensity is noncentrally maximal. Both types of imperfection sub-disk expand linearly in diameter with increase in index.

It is demonstrated briefly that certain dry collagen specimens exhibit small-angle diffraction in which the characteristics of the three types of sub-disk are apparent.

RECENT small-angle x-ray diffraction studies have disclosed that the collagen fibril (of kangaroo tail tendon) lacks the full complement of three-dimensional crystalline order, instead possessing periodic structure extending only along the fibril axis.[1] While x-rays are scattered from practically the whole cross section of the moist fibril, the fibril fractures longitudinally into thin filaments during drying, so that small-angle scatter is no longer coherent across its entire thickness.[2]

Simple, smooth, corrugated, and compound cylinder models for fibril structure[3] accounted for many of the gross features of collagen small-angle diffraction but were not satisfactory in detail.[2] The present paper considers rather more general models for distorted, one-dimensionally ordered diffractors. Involved is generalization of the previous compound cylinder model, whereby filament distortion is now assumed cylindrically symmetrical about any point in the fibril but periodically variable along its axis.

Since the models developed herein turn out to be, in fact, quite useful in the interpretation of collagen small-angle diffraction, the results are not of limited possibility of application in view of the widespread occurrence of collagen-type protein fibers.[4] It is not impossible that other systems of similar structure will be found, and while the present terminology is adopted particularly for convenience with reference to collagen, it is easily adapted to other situations. There is nothing in the development below which demands, for example, that the axial columns of material in the diffracting cylinder be actually preferentially connected into filaments.

THE IMPERFECT COMPOUND FIBRIL

To accomplish economy of development, the earlier discussion of Bear and Bolduan[3] will be assumed. Figure 1 shows what is meant by the imperfect compound fibril, contrasting longitudinal sections through related perfect and imperfect structures. The perfect fibril here considered is an ideal one with length L and radius R equal to that of the actual imperfect fibril. (The former is not, for example, the wet fibril of the collagen case, but a similar perfect structure in which length and radius are imagined to be contracted to dry-fibril values by removal of water *without* the normally accompanying distortions). Both fibrils are composed of thin cylindrical filaments of radius R_0, which throughout this discussion will be assumed so small relative to the total fibril radius (R) that the filaments form a continuum over fibril cross sections. Each per-

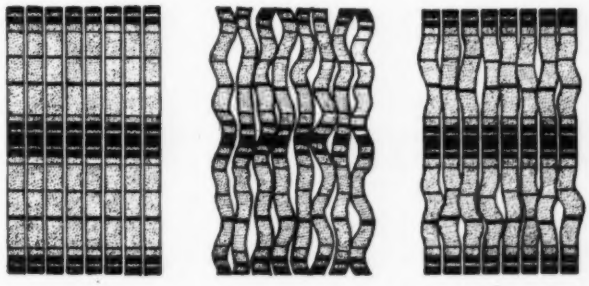


FIG. 1. Diagrammatic representation of portions of longitudinal sections through the perfect (a), imperfect (b), and mixed perfect and imperfect (c), fibrils considered in the text. Portions of eight filaments are shown running vertically in each type of fibril. The periodic pattern of bars and stippled areas running vertically along each filament indicates the characteristic features of longitudinal structure possessed alike by all. These features match transversely in (a), but because of distortions none of them do in (b), while only some do in (c).

* This paper represents a partial report on research sponsored by The Office of the Quartermaster General, Research and Development Branch, under Project No. 130-46 on "Determination of the nature and properties of skin structure" under direction of the Leather Subcommittee of the National Research Council Committee on Quartermaster Problems.

† Present address: Camp Detrick, Frederick, Maryland.

[1] O. E. A. Bolduan and R. S. Bear, J. Polymer Sci. **5**, 159 (1950).

[2] O. E. A. Bolduan and R. S. Bear, J. Polymer Sci. (to be published).

[3] R. S. Bear and O. E. A. Bolduan, Acta Crystallographica **3**, 236 (1950).

[4] Marks, Bear, and Blake, J. Exp. Zoöl. **111**, 55 (1949).

fect filament repeats its structure every b_0 distance along its axis.

Under these conditions the perfect fibril is essentially the previous smooth cylinder whose scattered intensity at the end of the vector ρ^* of reciprocal space may be written:

$$I_{p\rho^*} = I_0 i_e |F_{p\rho^*}|^2,$$

where I_0 is the incident intensity at the fibril and i_e is the scattering per electron for unit incident intensity at small angles. $F_{p\rho^*}$ is an integral over the cylindrical diffractor's volume:

$$\int_0^L N(y) \exp(i2\pi\bar{\nu} y y^*) dy \times \int_0^R \int_0^{2\pi} r \exp[i2\pi\bar{\nu} r r^* \cos(\psi - \psi^*)] dr d\psi,$$

where $N(y)$ is the density of electrons at level y, $\bar{\nu}$ is the wave number, $1/\lambda$, of the incident radiation, and r^*, ψ^*, and y^* are cylindrical coordinates for ρ^*. Note that $x = r\cos(\psi - \psi^*)$, the projection of r on the plane $\psi = \psi^*$, may be regarded as one of two orthogonal coordinates (x and z) in a cross section of the diffractor at level y. Figure 1 of Bear and Bolduan[3] illustrates the symbols employed.

It is now supposed that the elemental filament segments at r, ψ (or x, z), and y in the perfect fibril are displaced to r', ψ' (or x', z'), and y' in the imperfect fibril. The effect of the displacements is essentially that of altering the phases of the scattering from the segments. Since phase is influenced solely by the y' and x' coordinates, it is convenient to specify the corresponding displacements by $\Delta y = y' - y$ and $\Delta x = x' - x$. One can now express the total scattered intensity and the accompanying structure integral for the imperfect fibril just as for the perfect fibril, except that the primed coordinates must be used in the exponential factors. The primed coordinates can be eliminated by separating a factor,

$$\Phi = \exp[i2\pi\bar{\nu}(y^*\Delta y + r^*\Delta x)],$$

which is to be included in the structure integral, as expressed above, when it is adapted for the imperfect fibril.

Evaluation of the structure integral for the imperfect fibril, distinguished as $F_{q\rho^*}$, can proceed if the displacements can be expressed as functions of position in the fibril. Since these can be stated only statistically, determination of average scattered intensity is to be sought by calculation of an average $|F_{q\rho^*}|^2$. This involves integration twice through the volume of the diffractor, once as a volume element indicated by subscript m is carried through the cylinder, and again for an element labelled n, whose phase exponents are given negative signs.

It is now anticipated, as developed below, that the statistical formulation of the displacements depends on the radial separation, r_{mn}, between filaments passing through the mth and nth volume elements, which in general are at different levels, y_m and y_n. Since the diffractor is one-dimensionally (axially) ordered, the environment about any volume element at a given level will be statistically independent of position in that cross section, except as effects due to termination of structure at fibril boundaries enter. These latter difficulties are assumed negligible when distortion is sufficient to destroy rapidly the phase coherence of scattering by filament segments, as these become separated radially.

Under the above conditions, it becomes possible to prepare for subsequent integrations over cross sections by holding the mth volume element stationary at a given y_m and allowing the nth element to move over the entire cross section at y_n. The result multiplied by πR^2, the domain available to the mth element, summed also over the possible y levels, then yields the desired average $|F_{q\rho^*}|^2$:

$$\int_0^L \int_0^L N(y_m)N(y_n) \exp[i2\pi\bar{\nu}(y_m - y_n)y^*] dy_m dy_n \times \pi R^2 \int_0^\infty \int_0^{2\pi} r_{mn} \exp[i2\pi\bar{\nu} r_{mn} r^* \cos(\psi_{mn} - \psi^*)] \Phi_{mn} dr_{mn} d\psi_{mn}. \quad (1)$$

In this transformation the fact that

$$r_m \cos(\psi_m - \psi^*) - r_n \cos(\psi_n - \psi^*) = r_{mn} \cos(\psi_{mn} - \psi^*)$$

has been used, where ψ_{mn} is the angle between the original r_m and r_n. The integration over r_{mn} is extended to infinity for the reasons that are given above for neglecting boundary effects.

The quantity Φ_{mn} includes the effects of the displacements and will be an average value of an exponential factor which is

$$\exp\{i2\pi\bar{\nu}[y^*(\Delta y_m - \Delta y_n) + r^*(\Delta x_m - \Delta x_n)]\}. \quad (2)$$

It is a function of y_m, y_n, and r_{mn}, as will now be shown.

THE STATISTICS OF FILAMENT DISTORTION

In order to simplify this matter, it will be discussed as the general problem of the random displacement parallel to *any* line in the diffractor when the displacement of any pair of filament segments relative to each other is a function of the radial separation between the filaments.

The importance of the radial separation arises from the following basic assumptions. The forces of filament interaction tend to hold adjacent filaments parallel and with corresponding longitudinally distributed features of their structure in perfect register everywhere across fibril cross sections. Nevertheless, the filament interactions are not strong enough to prevent small displacements of filament segments from their ideal positions. Consequently, in passing from one filament segment to the corresponding segment of an adjacent filament, the latter experiences a small displacement, σ (root of a

mean square value), from its ideal position in a positive or negative manner along the given direction. Because of the one-dimensionally ordered distribution of matter in the system, it is expected that σ will possess one value, σ_b, for displacements along the axis of order (y), and another, σ_a, for any direction normal to the axis, such as the x direction. It is also postulated that both σ's are functions of the levels (y) at which the filament segments occur, since the chemical environment is also a function of level.

In passing across the fibril between two filaments of the volume elements m and n, separated by a distance r_{mn}, the chance of the nth filament's segment at a given level being displaced an appreciable amount, relative to the perfect position of the corresponding location on the mth filament, increases. The problem is similar to that of "random walk" along a line, where the quantity $r_{mn}/2R_0$, the number of filaments encountered in passing from m to n, each furnishing the possibility of the unitary displacement, corresponds to the "number of steps," each of σ-amount, taken in the positive or negative manner along the direction under consideration.

It is unnecessary to repeat the well-known mathematical treatment for the random walk problem. In the present instance it yields, for the probability that the total displacement along a general direction (u) shall be between Δu and $\Delta u+d\Delta u$, the gaussian distribution:

$$\left(\frac{R_0}{\pi r_{mn}\sigma^2}\right)^{\frac{1}{2}} \exp[-R_0(\Delta u)^2/r_{mn}\sigma^2]d\Delta u.$$

The average value for a displacement phase factor, such as $\exp(\pm i2\pi\bar{\nu}u^*\Delta u)$, where u^* is written as a general coordinate of reciprocal space, is then

$$\left(\frac{R_0}{\pi r_{mn}\sigma^2}\right)^{\frac{1}{2}} \int_{-\infty}^{+\infty} \exp\left\{\pm i2\pi\bar{\nu}u^*\Delta u - \frac{R_0(\Delta u)^2}{r_{mn}\sigma^2}\right\} d\Delta u$$
$$= \exp[-(\pi\bar{\nu}u^*\sigma)^2 r_{mn}/R_0]. \quad (3)$$

In the present evaluation of Φ_{mn}, note that the exponential term (2), given at the end of the last section, involves four displacements, Δy_m, Δy_n, Δx_m, and Δx_n. The probability that these are simultaneously experienced in two filament segments belonging to two filaments separated by r_{mn} is the product of four corresponding gaussian expressions with Δu suitably altered and σ properly subscripted. The averaging process, however, is separable into four distinct integrations, so that the final result is the product of four exponential terms such as resulted from integration (3), with two u^*'s and four different σ's suitably specified. Consequently, it is readily demonstrable that

$$\Phi_{mn} = \exp[-(A_m+A_n)r_{mn}/R_0], \quad (4)$$

where $A=(\pi\bar{\nu})^2[(y^*\sigma_b)^2+(r^*\sigma_a)^2]$. Note that A is a function of y because the σ's are also. The subscript on A indicates that the corresponding σ's are also to be given the same additional subscript.

THE INTENSITY SCATTERED BY THE IMPERFECT FIBRIL

It is now possible to return to the problem of evaluating average $|F_{q\rho^*}|^2$. As was indicated on the previous occasion for a similar integration in connection with the smooth cylinder, the present integration (1) with respect to ψ_{mn} yields $2\pi J_0(2\pi\bar{\nu}r_{mn}r^*)$. The integration with respect to r_{mn} then becomes

$$2\pi \int_0^\infty rJ_0(2\pi\bar{\nu}r^*r)\Phi_{mn}dr,$$

in which Φ_{mn} is given by Eq. (4) and subscripts on r have been dropped. Evaluated according to p. 386 of Watson[6] this becomes

$$\frac{2\pi R_0^2(A_m+A_n)}{[(A_m+A_n)^2+(2\pi\bar{\nu}R_0r^*)^2]^{\frac{1}{2}}} \equiv 2\pi R_0^2\Omega_{\rho^*}(mn), \quad (5)$$

where $\Omega_{\rho^*}(mn)$ is defined as shown.

Following procedures outlined previously, one can now determine that for the imperfect fibril, the intensity expression of the kth integrated disk of reciprocal space is:

$$I_{qkr^*} = I_0 i_e \frac{M\lambda}{b_0} 2\pi R^2 \pi R_0^2 \times \int_0^1 \int_0^1 \Omega_{kr}(mn)N(v_m)N(v_n) \times \exp[i2\pi k(v_m-v_n)]dv_m dv_n, \quad (6)$$

where $M\lambda/b_0$ is a factor introduced by projecting all of the kth disk's intensity onto its central plane at $y^*=k\lambda/b_0$, when M is the number of repetitions of structure along the length (L) of the fibril. A change of variable to $v=y/b_0$ has also been effected, and $\Omega_{kr^*}(mn)$ is the value of $\Omega_{\rho^*}(mn)$ at locations in reciprocal space now specified by k and r^*.

It is readily shown that Eq. (6) is least reliable, quantitatively, whenever somewhere in the structure A_m (or A_n) is zero, because then the infinite limit assumed during the radial integration is not justified and at $r^*=0$ the value of $\Omega(mn)$ incorrectly becomes infinite. Equation (6) is also inconvenient for general application to disk-shape studies because the terms of Ω depending on fibril structure (usually unknown *a priori*) and on radial location in reciprocal space are not separable. In view of these difficulties, it is generally necessary to consider special cases which may promise to be of interest. One of these was the compound fibril previously described,‡ and two more will be developed in the next two sections below.

[6] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Teddington, England, 1944).

‡ Bear and Bolduan (reference 3), in discussing a "compound fibril," stated, without proof, an average value for the square of a structure factor F''_{kr^*}. Filament displacements in this early model

THE IDEALLY IMPERFECT FIBRIL

For purposes of easy reference, it is convenient to describe an "ideally imperfect" fibril as one whose filaments are distorted, both axially and radially, by appreciable displacements occurring *everywhere* in the fibril. Under these conditions, the term $2\pi\tilde{\nu}R_0r^*$ occurring in the denominator of Eq. (5) may be neglected, particularly at small r^* and large k. Consequently, then

$$\Omega(mn) \doteqdot (A_m+A_n)^{-2}. \tag{7}$$

In the useful equations now to be developed it is also assumed that identical kinds of distortion are of nearly the same degree at all axial levels, though minor variations occur as a periodic function of level (v). In line with this assumption, the distortions may be written $\sigma_b^2=\sigma_{0b}^2+\epsilon_b$ and $\sigma_a^2=\sigma_{0a}^2+\epsilon_a$, where the ϵ's are departures from constant degrees of distortion expressed by the σ_0^2 terms. It is understood that the ϵ's alone are functions of v and they are regarded as small, relative to the squares of the σ_0's. It is also convenient to rescale r^* by letting $r^*=(k\lambda\sigma_{0b}/b_0\sigma_{0a})r^*_k$, where r^*_k is a new variable for the description of radial extension in reciprocal space.

It is then readily demonstrable that when $y^*=k\lambda/b_0$, to the approximation that the terms containing ϵ are small compared to $1+r^{*2}_k$,

$$\Omega_{kr^*}(mn) \doteqdot \frac{[1-(\epsilon_{bm}+\epsilon_{bn}/\sigma_{0b}^2)]+[1-(\epsilon_{am}+\epsilon_{an}/\sigma_{0a}^2)]r^{*2}_k}{4(k\pi\sigma_{0b}/b_0)^4(1+r^{*2}_k)^3}.$$

To a similar approximation,

$$1-\frac{\epsilon_m+\epsilon_n}{\sigma_0^2} \doteqdot \left(1-\frac{\epsilon_m}{\sigma_0^2}\right)\left(1-\frac{\epsilon_n}{\sigma_0^2}\right) \doteqdot \frac{\sigma_0^4}{\sigma_m^2\sigma_n^2},$$

since $1-(\epsilon/\sigma_0^2) \doteqdot \sigma_0^2/\sigma^2$. Consequently, one finally obtains

$$\Omega_{kr^*}(mn) \doteqdot \frac{(\sigma_{0b}^2/\sigma_{bm}\sigma_{bn})^2+(\sigma_{0a}^2/\sigma_{am}\sigma_{bn})^2r^{*2}_k}{4(k\pi\sigma_{0b}/b_0)^4(1+r^{*2}_k)^3}. \tag{8}$$

---

were solely axial and the same at all levels ($\sigma_a=0$, σ_b independent of y). Under these circumstances, since $A_m=A_n$ and both equal $(k\pi\sigma_b/b_0)^2$ at $y^*=k\lambda/b_0$, it is readily shown that

$$2\pi R_0^2\Omega_{kr^*}=\frac{\pi R_0^2(k\pi\sigma_b/b_0)^2}{2[(k\pi\sigma_b/b_0)^4+(\pi\tilde{\nu}R_0r^*)^2]^{\frac{1}{2}}}.$$

This, multiplied by the cross-sectional area of the fibril, πR^2, and divided by $(1/\pi R_0^2)^2$, because F''_{kr^*} as previously defined required a factor $(1/\pi R_0^2)$ expressing the number of filaments per unit of cross section, yields the average $|F''_{kr^*}|^2$ previously given. The filaments of the compound fibril were also assumed, though of much smaller radius than R, to be, nevertheless, of sufficient thickness to warrant retaining the Bessel function $[J_1(2\pi\tilde{\nu}R_0r^*)/\pi\tilde{\nu}R_0r^*]^2$ to express the effect of filament size on intensity distribution. The results of Bolduan and Bear (reference 2) indicate, however, that with collagen, R_0 is very small relative to the magnitudes involved at small diffraction angles, so that the present development has not included this term, which becomes unity at small R_0.

With this form for Ω, it is then possible to express Eq. (6) for the ideally imperfect fibril as follows:

$$I_{qkr^*} \doteqdot \frac{I_0i_eM\lambda\pi R^2\pi R_0^2(|F_{bk}|^2+|F_{ak}|^2r^{*2}_k)}{2b_0(k\pi\sigma_{0b}/b_0)^4(1+r^{*2}_k)^3}, \tag{9}$$

where

$$F_{bk} \equiv \int_q (\sigma_{0b}/\sigma_b)^2N(v)\exp(i2\pi kv)dv \tag{9a}$$

expresses the influence of the displacements in the direction of the y (or b) axis, and F_{ak} is similarly defined for the transverse (x or a axis) displacements with subscripts a written in place of b. These integrals receive greatest contributions to amplitude from intervals of v at which density (N) is large and displacement (σ) smaller than normal (σ_0), as is physically reasonable. That the integration extends over all v at which σ is appreciable, in this case the complete unit translation, is indicated by q below the integral sign.

THE MIXED PERFECT AND IMPERFECT FIBRIL

The ideally imperfect fibril is, by definition, one in which the displacements are never permitted to become zero at any level. On the other hand, the perfect fibril, in which displacements are everywhere zero, is essentially the previously described smooth cylinder, for which the equations analogous to (9) and (9a) are:

$$I_{pkr^*}=I_0i_e\frac{M\lambda}{b_0}(\pi R^2)^2\left[\frac{J_1(2\pi\tilde{\nu}Rr^*)}{\pi\tilde{\nu}Rr^*}\right]^2|F_{pk}|^2, \tag{10}$$

$$F_{pk} \equiv \int_p N(v)\exp(i2\pi kv)dv, \tag{10a}$$

wherein p indicates the perfection as did q the imperfection before.

Actual situations undoubtedly lie somewhere between the cases of Eq. (9) and (10), but the inadequacies of mathematical procedure make it difficult to obtain a single formulation capable of handling a system at whose various axial levels distortion may be lacking or assume appreciable degrees. It is possible, however, to consider a case in which over some segments of length the fibril is perfect, while at the remaining segments it is ideally imperfect, i.e., in short, the "mixed perfect and imperfect" fibril which is illustrated in Fig. 1c. The segments of length just mentioned are smaller than the structure period, b_0, of the fibril, which remains the distance along the axis over which the unit pattern of perfection and imperfection, as well as of filament density variation, extends.

An example of this sort would arise when there occur two essentially different types of filament interaction: one in which attraction between corresponding segments of adjacent fibrils is sufficient to hold them in perfect order, and another in which this attraction is weak, permitting development of appreciable filament distortion. That there are, in fact, two types of level,

bands and interbands, in collagen fibrils is readily seen in electron micrographs of stained preparation,[6] though the correspondence to the present regions of very different filament distortion is not immediately apparent, except as it seems generally true that fibril diameters are somewhat larger at the stained bands than at the unstained interbands (regions of perfection?). These facts are cited as *a priori* evidence of a possible value attached to the consideration of the mixed perfect and imperfect fibril.

It will be unnecessary to repeat very much of the mathematical treatment of this case, which follows fairly directly from what has been presented. Eventually, one finds for the intensity expression analogous to (6):

$$I_{skr^*}=I_{pkr^*}+I_{qkr^*}+I_{pqkr^*}. \tag{11}$$

The three terms will be considered in turn.

I_{pkr^*} is essentially as given in Eq. (10), except that the interval of F_{pk} is now less than the entire unit translation, being over the portion of v levels at which the structure is perfect. Consequently, this term expresses the intensity scattered purely by the perfect sections of the fibril.

I_{qkr^*} remains as stated in Eq. (9), but with the integration (9a) limited to v intervals over which the structure is ideally imperfect. This term indicates the intensity scattered solely by the imperfect segments of the fibril.

I_{pqkr^*} is a hitherto undiscovered term involving the contributions to scattering resulting from interference between the perfect and imperfect levels of the fibril. It may be determined by returning to the integral (1), which is to be modified as follows: one first lets m designate intervals of y level at which structure is ideally imperfect, and n those where perfection exists; the integration limits are, therefore, to be suitably restricted. It also follows that the Φ-factor is a function only of displacements at the m levels, since Δy_n and Δx_n are zero in (2). The evaluation of average Φ and the integrations proceed just as before. At the state (7), Ω becomes $\Omega(m)\doteqdot A_m^{-2}$, which may, according to the same approximations as before, be written

$$\Omega_{kr^*}(m)\doteqdot\frac{(\sigma_{0b}/\sigma_{bm})^4+(\sigma_{0a}/\sigma_{am})^4r^{*}{}_k^2}{(k\pi\sigma_{0b}/b_0)^4(1+r^{*}{}_k^2)^3},$$

which is analogous to Eq. (8).

A similar development, with m and n interchanged, is also required, so that one finally obtains, corresponding to (9):

$$I_{pqkr^*}\doteqdot\frac{2I_0i_eM\lambda\pi R^2\pi R_0^2}{b_0(k\pi\sigma_{0b}/b_0)^4}\times\frac{(G_{bk}F^*_{pk}+F_{pk}G^*_{bk})+(G_{ak}F^*_{pk}+F_{pk}G^*_{ak})r^{*}{}_k^2}{(1+r^{*}{}_k^2)^3}, \tag{12}$$

[6] F. O. Schmitt and J. Gross, J. Am. Leather Chemists' Assoc. 43, 658 (1948).

where the new structure integrals are

$$G_{bk}\equiv\int_q(\sigma_{0b}/\sigma_b)^4N(v)\exp(i2\pi kv)dv \tag{12a}$$

and G_{ak}, which is similarly defined with subscripts b changed to a.

DISTRIBUTION OF INTENSITY OVER RECIPROCAL-SPACE DISKS

The primary aim of this investigation is to indicate how the presence of perfection or imperfection may be recognized from observed distributions of diffracted intensity over disks of reciprocal space. For this purpose, it is convenient to rewrite Eq. (11) by use of (9), (10), and (12):

$$I_{skr^*}=I_0'\left\{U_k\left[\frac{J_1(2\pi\tilde{\nu}Rr^*)}{\pi\tilde{\nu}Rr^*}\right]^2+\frac{V_k}{(1+r^{*}{}_k^2)^3}+\frac{W_kr^{*}{}_k^2}{(1+r^{*}{}_k^2)^3}\right\}, \tag{13}$$

where

$$I_0'\equiv I_0i_eM\lambda\pi R^2/b_0,$$

$$U_k\equiv\pi R^2|F_{pk}|^2,$$

$$V_k\equiv\frac{\pi R_0^2}{2(k\pi\sigma_{0b}/b_0)^4}[|F_{bk}|^2+4(G_{bk}F^*_{pk}+F_{pk}G^*_{bk})],$$

$$W_k\equiv\frac{\pi R_0^2}{2(k\pi\sigma_{0b}/b_0)^4}[|F_{ak}|^2+4(G_{ak}F^*_{pk}+F_{pk}G^*_{ak})],$$

$$r^{*}{}_k\equiv(b_0\sigma_{0a}/k\lambda\sigma_{0b})r^*.$$

The significance of Eq. (13) and the auxiliary relations is difficult to assess without reference to the relative magnitudes of the various quantities, many of which are unknown *a priori* in an experimental investigation. In general, one may expect for moderate and large values of k that the coefficients of the structure factors in V_k and W_k will be considerably smaller than the corresponding coefficient in U_k, since it has been assumed above that $R_0\ll R$ and that the results are not to be applied for excessively small σ_{0b}. Indeed, because of this fact, one may immediately note that, except when the F_{pk} structure factors are small enough in amplitude, it will be difficult to detect evidences of imperfection because the first term of Eq. (13) will predominate. Accordingly, the system will often act approximately as the smooth cylinder previously discussed.

On the other hand, consider the cases in which the $|F_{pk}|$ are generally smaller than the structure factors involving imperfection, but not entirely negligible, being still sufficient to cause U_k to be comparable in magnitude to V_k and W_k. In this situation the products of the type $|F_{pk}||G_k|$ are expected to be unimportant relative to $|F_{bk}|^2$ or $|F_{ak}|^2$. Then one knows *a priori* only that the terms U_k, V_k, and W_k of Eq. (13) may

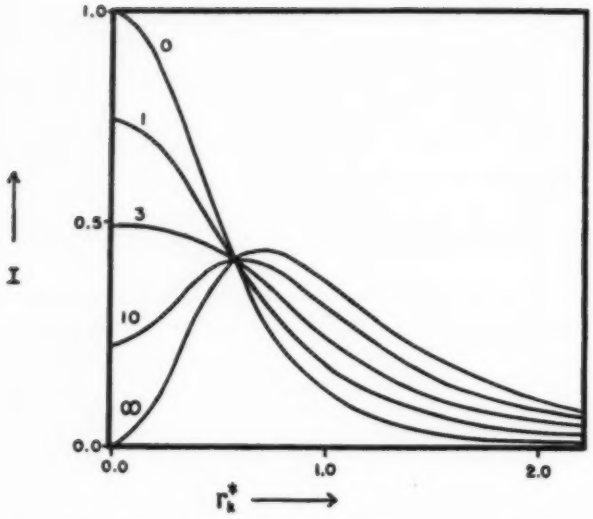


FIG. 2. The range of shapes available to the combined sub-disks of imperfection at a given kth level in reciprocal space. Ordinates represent intensity in arbitrary units encountered as one proceeds from the disk center along radial distances expressed by r^*_k. The figures accompanying each curve (their order is reversed on the right of the common intersection) indicate the ratio of the weights given to the sub-disks of radial and longitudinal imperfection (W_k/V_k) in obtaining each characteristic shape. Ordinates have been scaled for each curve so that all have equal areas under them.

have various positive real values. Furthermore, a value is not immediately apparent for the scaling factor relating r^* and r^*_k, which, respectively, govern how intensity from perfect regions (first term of Eq. 13) and from imperfections (second and third terms) may be spread from the center of the disk. Because of these facts, one may expect manifestations of perfection and imperfection to be in effect independent. Indeed, one may speak of each reciprocal space disk as being composed of superimposed sub-disks of perfection and imperfection. The following paragraphs state the properties of these types of sub-disk, indicating how they may be distinguished.

Sub-Disks of Perfection

These have exactly the shape and diameter properties following from the Bessel-function multiplier of U_k, as previously described for the smooth cylinder. It will be unnecessary to discuss them again, except to recall that disk diameters are independent of index k and reciprocal to R.

Sub-Disks of Longitudinal Imperfection

While experimentally both types of imperfection scatter intensity over the total kth disk as a function of r^*_k (hence, may be difficult to distinguish entirely), it is of interest to consider the effects of longitudinal and radial imperfection separately, since they appear independently in Eq. (13). The shape factor associated with V_k is $(1+r^{*2}_k)^{-2}$, which has value unity at r^*_k (or r^*) zero, and declines symmetrically away from this center. Thus, the sub-disks of longitudinal imperfection and of perfection have similar shapes and central maxima. They tend to merge together into a combination of intermediate diameter. However, because r^*_k is a function of index k, the diameter of the imperfection sub-disk increases proportional to that index. At high indices, the longitudinal imperfection sub-disk tends to produce a declining shoulder extending beyond the sub-disk of perfection.

Sub-Disks of Radial Imperfection

The factor associated with W_k in Eq. (13) furnishes a shape function for this sub-disk. At the center (r^*_k or r^* zero) it has no intensity but reaches a maximum value at $r^*_k=1/\sqrt{2}$, after which it again declines to zero at large r^*_k. Because r^*_k is reciprocal to k, the observed r^* at which the maximum occurs becomes larger as k increases, and the maximum may eventually appear outside the sub-disk of perfection at large k. The distinctive maximum cannot, however, escape the sub-disk of longitudinal imperfection, except at k values where W_k is sufficiently larger than V_k that the longitudinal sub-disk is relatively faint. The range of various possible shapes permitted for the combination of the two types of sub-disk of imperfection is shown in Fig. 2.

RESULTS WITH COLLAGEN

The various sub-disk properties are nicely portrayed in small-angle diffraction patterns obtained under certain conditions from collagen specimens. Figure 3 shows one such pattern, obtained with a previously described pinhole camera[7] from plain dry kangaroo tendon. Further particulars concerning the experimental conditions, and detailed information regarding such patterns, will be given elsewhere. At present it is sufficient to point out their principal features.

The general expansion of layer lines with increase in index is outstandingly apparent. According to the theory of the sub-disks of imperfection, to which the expansion is due, this should be linear with k, as is observed. In the example given, beyond about the fifth layer line the spreading is sufficient to permit identification of various features of individual types of sub-disks. Thus, the fifth and seventh lines are largely the centrally intensified, short lines characteristic of sections through sub-disks of perfection. The eighth and eleventh lines possess, in addition to central maxima corresponding largely to perfection sub-disks, the definite lateral intensity maxima expected of sub-disks of radial imperfection. The ninth line appears to be shaped predominantly like a sub-disk of longitudinal imperfection, with perhaps some central contribution from perfection. The strong sixth layer line seems to have all types of sub-disk well represented, since in-

[7] O. E. A. Bolduan and R. S. Bear, J. Appl. Phys. 20, 983 (1949).

tensity is maintained at a high level all along its course. Similar layer-line spreading and structure can be faintly observed over the remainder of the first score of diffractions, but become difficult to observe beyond this.

Estimation of the parameters of fibril size (R) and of filament radius (R_0) and displacement (σ_0's) can be made only in approximate fashion for the present case. Bolduan and Bear[2] found the effective fibril radius for diffraction to the first layer line (for which imperfections do not materially influence the coherence of diffraction from the entire fibril cross section) to be of the order of 1000A, a reasonable estimate for R. The evidence relative to the earlier and simpler compound fibrillar model suggested that R_0 is very small, and from the fact that the largest equatorially oriented spacing of dry collagen is about 10A, the filament radius may be taken as that appropriate for a single polypeptide chain ($R_0 \doteqdot 5.0$A).

The present collagen diagram permits calculation of the ratio σ_{0a}/σ_{0b}, which is about 1.85 according to the observed positions of the maxima on the eighth and eleventh layer lines. The maxima are expected at $r^*_k = 1/(2)^{\frac{1}{2}}$, according to the simple theory of Eq. (13). However, the neglect, in passing between Eqs. (5) and (7), of the term $2\pi\bar{\nu}R_0r^*$, may not be justified at the appreciable r^* of the intensity maxima, unless compensatory additional weight is placed on the σ_a terms of A_m and A_n. Thus, the calculated ratio σ_{0a}/σ_{0b} may be excessive and it may be concluded at the moment only that σ_{0b} and σ_{0a} are of comparable magnitude.

It may reasonably be expected that the σ_0's are of the order of size of R_0. With this assumption, and using the estimates of R and R_0 given above, along with $b_0 = 627$A, appropriate to the present example, one can readily justify the remarks regarding the relative magnitudes of the coefficients in U_k, V_k, or W_k, and the structure factors of Eq. (13), *providing* one does not attempt to make application to the first four layer lines.

From the above statements it is clear that the approximations applied in the theoretical development are least satisfactory at small k and large r^*. With regard to the latter variable, it seems likely that the relatively sharp terminations of the individual layer lines evident in Fig. 3 may not be adequately accounted for at present. It is difficult, in the theoretical treatment, to allow for the fact that filament segments cannot be displaced with utter disregard for the space-filling properties of adjacent filaments, and correction for this limitation might explain the observed abrupt line terminations. Nevertheless, nothing was neglected in the present theory which may be expected to influence markedly the general kind of distribution of intensity in reciprocal space. The ability of the theory to account for the general observations with collagen offers a fairly conclusive demonstration of the appropriateness of the mixed perfect and imperfect model for the structure of dry collagen fibrils.

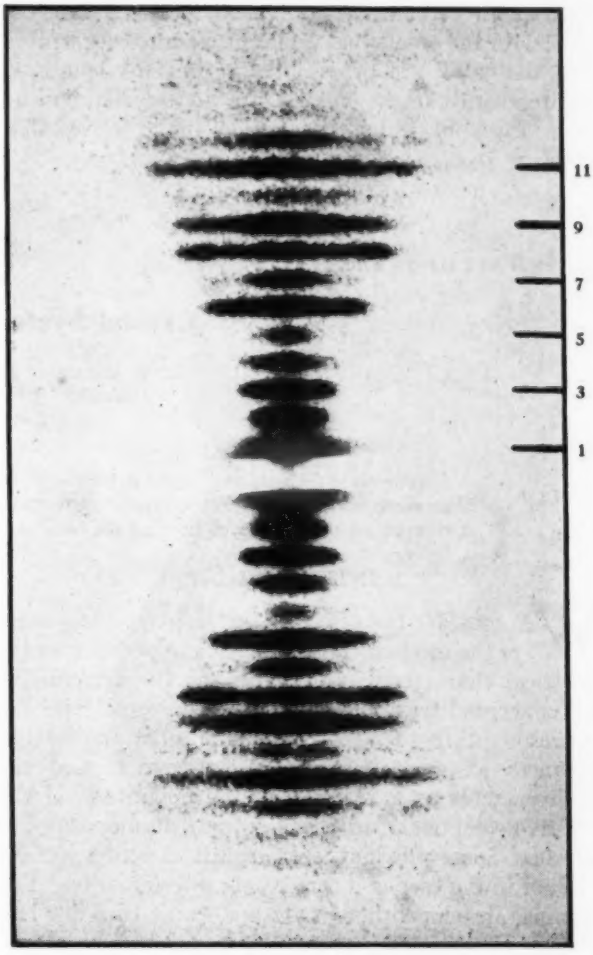


FIG. 3. A pinhole diffraction pattern obtained from a plain dry fiber of kangaroo tail tendon at perpendicular incidence of a Cu Kα (Ni-filtered) beam. Specimen-to-film distance 15 cm, exposure 156 hr at 35 kvp, 20 ma. Camera resolution 800A vertically (along fiber axis), 400A horizontally. Fundamental period of sample: $b_0 = 627$A. The odd indices (k) of the layer lines are indicated.

The present development has determined shape properties of reciprocal space disks in much more detailed fashion than did the previous consideration of simpler models for one-dimensionally order diffractors. The earlier applications, in terms of the effective diffractor radii (R_k), are useful when distortion is insufficient to produce distinguishable imperfection sub-disks, or when one is making preliminary rough examinations to judge the types of distortion apparent in a given case. The present detailed shape considerations then permit further examination if the diffraction data warrant it.

In all of the models which have been discussed for one-dimensional diffractors, the distortions from the perfect, smoothly cylindrical case are ones which affect only spreading of diffraction along layer-line lengths and not broadening of these lines in the axial direction. While axial *filament* distortions have been assumed in some cases, these have been taken as statistically iden-

tical at all periodically equivalent locations along the axis, with the result that adjacent repeated segments of the *total fibril* are of essentially constant length, b_0. Consequently, these models are to be distinguished from cases in which the period may be variable.[8]

[8] See R. Hosemann, Z. Physik **127**, 16 (1949).

Small-angle diffraction data on collagen have not as yet been obtained at sufficient angular resolution to demonstrate that this sort of fibrillar axial distortion is present, but it is in any case of much less importance than the intra-fibrillar filament distortions which produce the readily observed variations in line length.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

Crystal Systems with Low Loss

RUTH BAUMANN FRY AND WILLIAM J. FRY
University of Illinois, Urbana, Illinois
(Received August 25, 1950)

Approximate formulas are derived for computing the effect of mechanical resistance on the electrical characteristics of a piezoelectric crystal system for the case of a general reactive termination and low loss. A comparison of values computed on the basis of these formulas and exact calculations is presented.

I. INTRODUCTION

WE consider the effect of the resistive component of the mechanical terminating impedance on the electrical characteristics of a piezoelectric system. We are concerned with the shift of the resonant and the antiresonant frequencies, the magnitudes of the impedances at resonance and antiresonance, and the quality factor as a function of the magnitude of the resistive component of the mechanical terminating impedance. Some previous work in this direction was restricted to the case of a zero reactive termination.[1] The formulas developed herein are applicable to a low loss system vibrating in either longitudinal or thickness mode in which the reactive component is unrestricted.

II. ANALYSIS

Table of Symbols

A_c Area of radiating face of crystal.
C $\text{Cos}\gamma_0$.
C_0 Clamped capacity of crystal.
d Piezoelectric constant.
f Frequency.
f_r Resonant frequency.
Δf Difference between resonant and antiresonant frequencies.
K V_c/L_c.
l_w Width of crystal.
L_c Length of crystal.
M $\gamma_0\theta$.
Q' Quality factor, ratio of the impedance at antiresonance and the impedance at resonance.
s Elastic constant.
S $\text{Sin}\gamma_0$.
V_c Velocity of sound in crystal.
x Coordinate variable.
X_a+jY_a Z_{m2}/Z_0.
Y_a See definition of X_a.
Z_e Input electrical impedance into crystal.
Z_0 $A_c\rho_c V_c$.
Z_{m2} Mechanical impedance of load on crystal.
Z_{ar} Electrical impedance of crystal system at antiresonance.
Z_r Electrical impedance of crystal system at resonance.

Greek Symbols

γ_c $\omega L_c/V_c$.
γ_0 Value of γ_c at resonance for lossless system.
ϵ $\gamma_0-\gamma_c$.
θ $\varphi^2/Z_0K_0C_0\gamma_0$.
μ $[Y_a^2+1+(dY_a/d\gamma)_{\gamma_0}]\epsilon$.
ρ_c Density of crystal.
φ $=dl_w/s$ for longitudinal mode.
$=dA_c/sL_c$ for thickness mode.
d and s in these formulas are not identical in value for the two modes.
ω $2\pi f$.

Consider a crystal vibrating in longitudinal or thickness mode and free at one end, $x=0$, but with a load both resistive and reactive on the opposite end, $x=L_c$

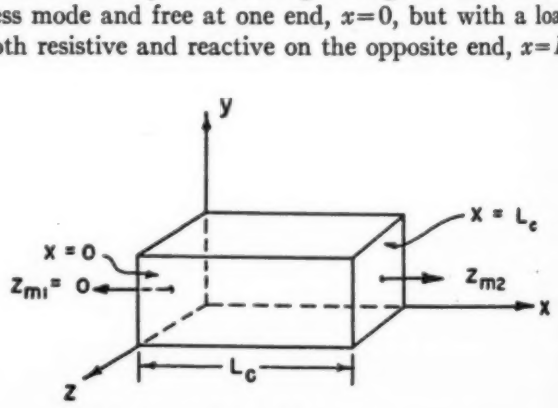


FIG. 1. A crystal placed in a coordinate system.

[1] W. J. Fry, J. Acoust. Soc. Am. **21**, 29 (1949).

(see Fig. 1). The general expression for the electrical input impedance under these conditions is[2]

$$Z_e=\frac{1}{j\gamma_e KC_0}+\frac{\varphi^2}{Z_0}\left[\frac{-2(1-\cos\gamma_e)+(jX_a-Y_a)\sin\gamma_e}{(X_a+jY_a)\cos\gamma_e+j\sin\gamma_e}\right]. \quad (1)$$

The resonant and antiresonant frequencies are defined as those frequencies for which the absolute value of the electrical input impedance has its minimum and maximum values.

In (1) insert the relation $\gamma_e=\gamma_0-\epsilon$ and the condition for resonance of a no loss system, $Y_a\cos\gamma_0+\sin\gamma_0=0$, represent $\cos\gamma_e$ by $\cos(\gamma_0-\epsilon)$ and $\sin\gamma_e$ by $\sin(\gamma_0-\epsilon)$; approximate these by $\cos\gamma_0(1-\epsilon^2/2)+\epsilon\sin\gamma_0$ and $\sin\gamma_0(1-\epsilon^2/2)-\epsilon\cos\gamma_0$, respectively; also let

$$Y_a=Y_{a0}-\epsilon(dY_a/d\gamma)_{\gamma 0}+\tfrac{1}{2}\epsilon^2\left(\frac{d^2Y_a}{d\gamma^2}\right)_{\gamma_0}.$$

Form the quantity $|Z_e|^2$. The highest power of ϵ retained is the second. To obtain the values of ϵ for which $|Z_e|^2$ has its maximum and minimum, the expression for $|Z_e|^2$ is differentiated with respect to ϵ and the result equated to zero. Powers up to and including the fifth power of ϵ are retained in the expression; however, all powers in X_a higher than the lowest in each coefficient are rejected, since the range for X_a is $0\leq X_a\leq 0.05$. Hence, the form

$$\left\{\begin{aligned}&[(1-C)^2C^4\beta^6+\theta(1-C)^2C^2\beta^3E]\epsilon^4\\&+[\theta(1-C)^4C^2\beta^4]\epsilon^3\\&+[3(1-C)^2C^2\beta^3X_a^2SC\theta\\&\qquad-3(1-C)^2C^2\beta EX_a^2\theta\\&\qquad\quad+6(1-C)^3C^3\beta^2Y_aX_a^2\theta]\epsilon^2\\&+[(1-C)^4C^2\beta^2X_a^2\theta]\epsilon\\&+[2(1-C)^4X_a^4SC\theta-(1-C)^2C^4\beta X_a^4\\&-(1-C)^2C^3\beta X_a^4S\theta-2(1-C)^3C^3Y_aX_a^4\theta]\end{aligned}\right\}=0, \quad (2)$$

where

$$\beta=Y_a^2+1+(dY_a/d\gamma)_{\gamma_0},$$

and

$$E=(C-1)^2[-2Y_a(dY_a/d\gamma)_{\gamma_0}-\tfrac{1}{2}(d^2Y_a/d\gamma^2)_{\gamma_0}]-Y_a[1+(dY_a/d\gamma)_{\gamma_0}C]^2.$$

Although these coefficients appear unwieldy and irreducible, an approximation of the form

$$\mu^4+\theta\left(\frac{1-C}{C}\right)^2\mu^3+\left(\frac{1-C}{C}\right)^2X_a^2\theta\mu-X_a^4=0, \quad (3)$$

gives a result which is accurate to within ten percent for the quantities of interest when Y_a is on the range $0\leq Y_a\leq 2$.

The real roots of this equation correspond to the resonant and antiresonant frequencies. Explicitly,

$$\mu_r=-\left(\frac{C-1}{C}\right)^2\frac{\theta}{2}+\left(\frac{C-1}{C}\right)^2\frac{\theta}{2}\left[1+X_a^2\Big/\left(\frac{1-C}{C}\right)^4\left(\frac{\theta}{2}\right)^2\right]^{\frac{1}{2}}, \quad (4a)$$

$$\mu_{ar}=-\left(\frac{C-1}{C}\right)^2\frac{\theta}{2}-\left(\frac{C-1}{C}\right)^2\frac{\theta}{2}\left[1+X_a^2\Big/\left(\frac{1-C}{C}\right)^4\left(\frac{\theta}{2}\right)^2\right]^{\frac{1}{2}}. \quad (4b)$$

From these results, it follows that

$$\frac{\Delta f}{f_r}=\frac{\{[(1-\cos\gamma_0)/\cos\gamma_0]^4\theta^2+4X_a^2\}^{\frac{1}{2}}}{[Y_a^2+1+(dY_a/d\gamma)_{\gamma_0}]\tan^{-1}(-Y_a)}. \quad (5)$$

By evaluating the electrical input impedance at the resonant and antiresonant frequencies, a quality factor

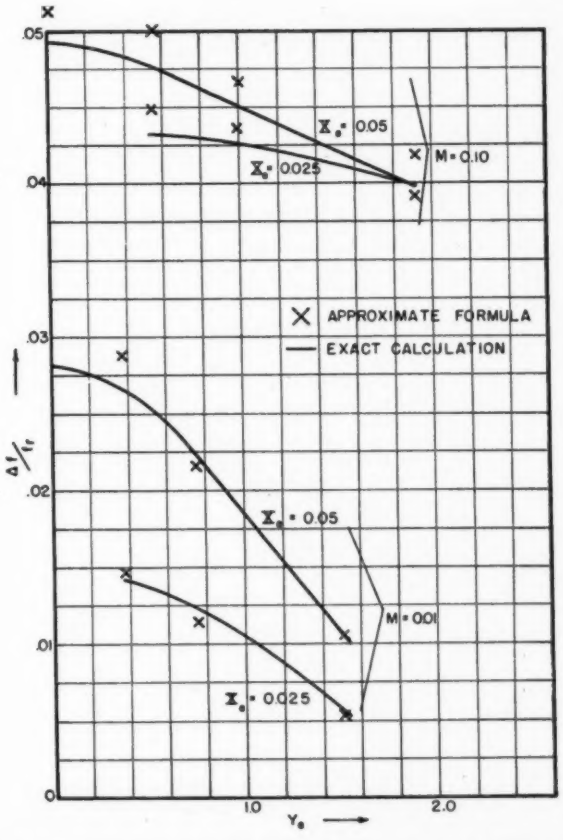


FIG. 2. Comparison of exact calculations with results computed from the approximate formula for the quantity $\Delta f/f_r$.

[2] See Fry, Taylor, and Henvis, *Design of Crystal Vibrating Systems* (Dover Publications, New York, 1948), for one derivation.

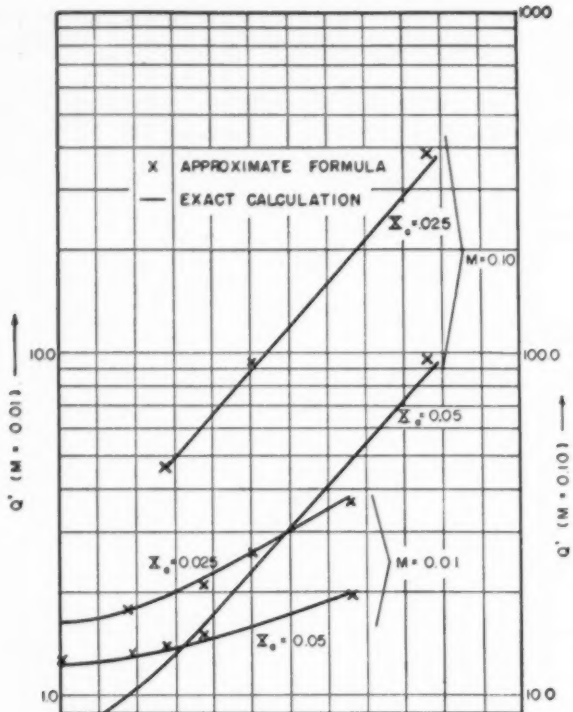


FIG. 3. Comparison of exact calculations with results computed from the approximate formula for the quality factor Q'.

is obtained. This quantity is defined as the absolute value of the ratio of the impedance at antiresonance to the impedance at resonance. Using the values found for μ_{ar} and μ_r given by expressions (4a) and (4b), and the equation representing Z_e, making reductions of form similar to those made to obtain μ_r, μ_{ar} the representations for $|Z|_r$ and $|Z|_{ar}$ follow:

$$|Z_r| = \frac{([\cos\gamma_0 - 1]/\cos\gamma_0)^2 \tfrac{1}{2}\theta}{X_a K C_0 \gamma_0} \times \left\{ -1 + \left[1 + \frac{X_a^2}{([\cos\gamma_0 - 1]/\cos\gamma_0)^4 (\tfrac{1}{2}\theta)^2} \right]^{\frac{1}{2}} \right\}, \quad (6a)$$

$$|Z_{ar}| = \frac{([\cos\gamma_0 - 1]/\cos\gamma_0)^2 \tfrac{1}{2}\theta}{X_a K C_0 \gamma_0} \times \left\{ 1 + \left[1 + \frac{X_a^2}{([\cos\gamma_0 - 1]/\cos\gamma_0)^4 (\tfrac{1}{2}\theta)^2} \right]^{\frac{1}{2}} \right\}. \quad (6b)$$

The quality factor, $Q' = |Z_{ar}/Z_r|$, is then obtained directly as

$$Q' = \frac{[1 + X_a^2/\{([\cos\gamma_0 - 1]/\cos\gamma_0)^4 (\tfrac{1}{2}\theta)^2\}]^{\frac{1}{2}} + 1}{[1 + X_a^2/\{([\cos\gamma_0 - 1]/\cos\gamma_0)^4 (\tfrac{1}{2}\theta)^2\}]^{\frac{1}{2}} - 1}, \quad (7)$$

which is seen to be the ratio of μ_{ar} to μ_r.

For $Y_a = 0$ and $0 \leq X_a \leq 0.05$, the quantities $\Delta f/f_r$, $|Z_r|$, $|Z_{ar}|$, and Q' reduce to those given in a previous paper.[1]

For a crystal system, expressions (5) and (7) relate the quality factor, Q', and the quantity $\Delta f/f_r$ to the quantity X_a, which is the ratio of the acoustic input resistance into the backing to the characteristic impedance of the crystal. This quantity, X_a, characterizes the losses.

III. COMPARISONS

Because of the large number of steps involved in the derivation in which approximations were introduced, it is somewhat difficult to specify the ranges of applicability of these expressions for a given accuracy in Q' and $\Delta f/f_r$ without some numerical comparisons. Accordingly, a numerical comparison was carried out between the values computed on the basis of these formulas and on the basis of the exact relation (1) from which these expressions were derived. The results of this work appear in Figs. 2 and 3. The specific examples chosen correspond to a quartz mercury system, $M = 0.01$, and to an ADP mercury system, $M = 0.1$. Calculations were made for four lengths of mercury. The value of Y_a is related to the percentage frequency shift expressed as a percent of the resonant frequency of the free crystal. For the ADP Hg system the maximum shift for which values were computed is 35 percent and for the quartz Hg system 32 percent. Figure 2 indicates the results for $\Delta f/f_r$ and Fig. 3 illustrates the results on the quality factor Q'. It should be pointed out that it was not the purpose of this numerical work to obtain accurate numerical estimates of the difference between exact calculations and those using the expressions (5) and (7) for the purpose of adding correction terms, but to obtain only an estimate of the percentage error involved for a range of the variables X_a and Y_a. As the result of this analysis, we can state that expressions (5) and (7) can be used for the region of the variables, Y_a and X_a, $0 \leq Y_a \leq 2$ and $0 \leq X_a \leq 0.05$ with an error not exceeding roughly 12 percent in the value $Q' - 1$ and an error not exceeding 10 percent in the quantity $\Delta f/f_r$. As X_a approaches zero, the error approaches zero, so that if X_a is further restricted, values calculated on the basis of these formulas are correspondingly more accurate.

Density Measurements in Supersonic Flow by Means of the X-Ray Absorption Method*

EVA M. WINKLER
U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland
(Received August 31, 1950)

The density in supersonic flow was measured by the absorption of soft x-rays. The equipment is briefly described. Data obtained in undisturbed flow at $M=2.83$ are compared with values computed from Pitot pressure measurements. Measurements of the density distribution across a shock wave on a wedge-shaped model are presented. The suitability of the method, its limitations, and improvements of the technique are briefly discussed.

THE x-ray absorption method was used to measure the air density distribution in the free stream of a supersonic nozzle and along a traverse through an oblique shock wave. The measurements were made in the NOL 18×18 cm continuous wind tunnel at a Mach number of 2.83.

The equipment incorporates a Machlett AEG 50 x-ray tube with a 1-mm beryllium window. The x-ray intensity is measured by Geiger-Mueller counters (type TGC-1). The x-ray tube and one counter are mounted to their respective tunnel window panels. By means of a Y-shaped connection tube, between x-ray tube and tunnel window, a second counter is mounted close to the x-ray tube. It receives x-rays from about the same target area as does the measuring counter and is utilized to determine the "zero" x-ray intensity. The connection tube is evacuated to avoid unnecessary absorption. The ray entrance and exit pinholes in the tunnel windows are sealed by 20μ-Cellophane foils. An x-ray beam averaging 1.5 mm in diameter was used throughout all measurements. The x-ray tube was operated at 4 kv, which corresponds to an average effective wavelength of 3.7A.

The absorption of 4-kv x-rays in an 18-cm air path of 0.1 atmospheric density, which is established in the test section at $M=2.83$ having atmospheric supply air, is about 20 percent. Accordingly, the density measurements in the undisturbed flow are expected to be accurate to within ± 4.5 percent if the x-ray intensity is measured with ± 1.2 percent as it was during the tests.

Two series of measurements were made along the center line of the tunnel test section to obtain a density survey of the undisturbed flow and two other series along a traverse through a shock wave. For each series the individual points were measured at least 10 times. Each measurement required about 12 sec.

The results of the density survey are given in Fig. 1. The scattering of the individual measurements around the mean values amounts to between 3 and 6 percent. The comparison with the Pitot pressure survey is given in Figs. 1(B) and 1(C).

The objective of the measurements presented in the following was primarily to get orientating informations if this technique is to be applied to the study of flow patterns. A wedge-shaped model was available. It extended almost the full span across the test section (test section width, 182 mm; model span 178 mm). A sche-

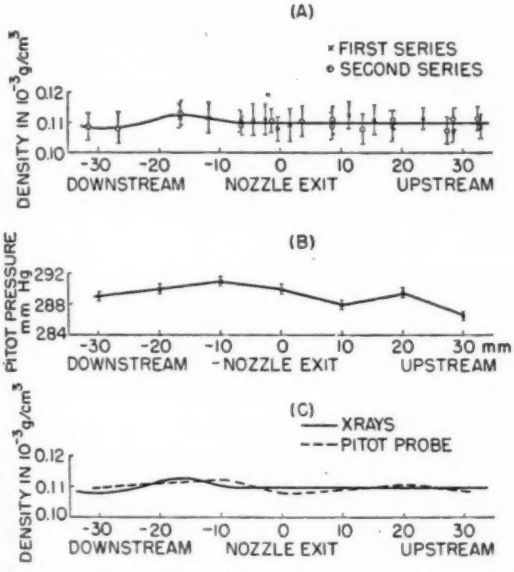


FIG. 1. Air density survey, $M=2.83$. (A) X-ray measurements. (B) Pitot pressure measurements. (C) Comparison of x-ray measurements and density computed from Pitot pressure measurements.

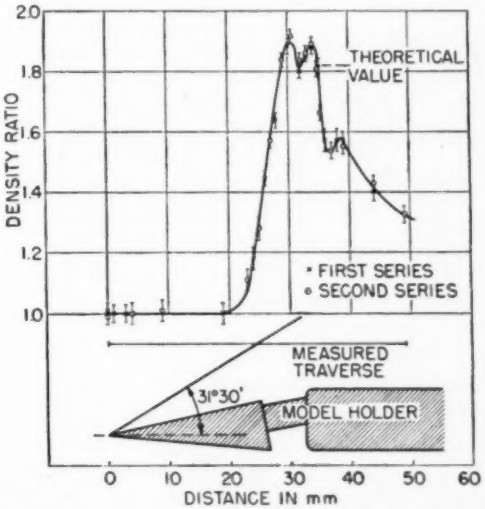


FIG. 2. X-ray measurements of the density distribution of a shock wave; $M=2.83$.

* First measurements of this kind have been made by Arnold in the Kochel 40×40 cm supersonic wind tunnel in 1944 and 1945.

matic diagram of the model, the shock wave, and the respective location of the measured traverse is given in Fig. 2 together with the results of the measurements.

The general appearance of the measured density distribution compares with schlieren pictures taken of the flow pattern. The schlieren picture also indicates a double maximum after the density rise which is followed by an expansion wave. The maximum at 38 mm from the leading edge can be correlated with a compression wave originating at the model holder.

The extension of the density rise over a distance of about 7 mm cannot be explained completely by the finite size of the x-ray beam and the fact that the traverse intersects the shock wave at an angle of 31.5°. The total density increase across the shock overshoots the increase computed from shock angle and Mach number by 4.5 percent. This overshoot cannot be accounted for by the scattering of the individual measurements, which is only 1.5 or 2 percent in that portion of the curve.

As the measurements of the undisturbed flow illustrate, the x-ray method is well suited for nozzle calibration. The advantage that the amount of absorption can be adjusted suitable to the density range to be measured, by properly selecting the wavelength, makes this technique applicable to a wide range of densities. Limitations, however, are imposed by the technical difficulties in minimizing the absorption in windows, especially the x-ray tube window, if for decreasing density values, wavelengths of about 8A and longer need to be employed. The study of flow patterns requires smaller beam diameters and higher x-ray intensity than were used for the orientating measurements given in Fig. 2. In comparison to optical methods refraction and deflection phenomena can practically be neglected with x-rays. This is of advantage for measurements near wall and model surfaces or measurements of steep density gradients.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

Deceleration and Ionizing Efficiency of Radar Meteors

D. W. R. McKINLEY
National Research Council, Ottawa, Canada
(Received September 5, 1950)

Improved instrumentation of the radar system has permitted a more accurate analysis of range-time records of meteor echoes and enabled the meteor deceleration to be determined in a few special cases. Three illustrative examples are described in which the measured mean decelerations were -0.48, -1.1, and -1.5 km sec^{-2}. Velocity data from a continuous wave Doppler system were also available for these meteors and are compared with the radar data.

Using existing atmospheric density values and Lovell's scattering formula, it is computed that the ionizing efficiency, or the fraction of the kinetic energy of the meteor converted to ionization, is 10^{-6} for a 60-km/sec meteor and 10^{-8} for a 20-km/sec meteor. If lower values of the air density are adopted, the efficiency figures are increased and at the same time the gap is narrowed between the rates of electron production deduced from loss of mass considerations and the rates calculated from the radar data. More statistical data on radar meteors and further information on the ionizing properties of 1000 electron-volt atoms are required before definite conclusions can be drawn.

I. INTRODUCTION

IN a previous paper[1] several methods were described for reducing radar range-time records of meteors to yield the velocity of the meteor from a single station observation. Both the velocity and the path position could be found if data were available from three stations. In essence these methods involve fitting to the photographic record of the meteor echo a theoretical curve of the form,

$$R^2=R_0^2+v^2(t-t_0)^2, \qquad (1)$$

where R is the range to the target at time t, R_0 is the minimum range at time t_0, and v is the meteor velocity.

In this work it was assumed, first, that the meteor path was a straight line, and second, that the meteor velocity was constant throughout the observed path. The attraction of the earth will cause a downward deviation of the meteor which is negligibly small. For example, the longest duration of an echo moving with the velocity of the meteor (h characteristic), that we have observed was ten seconds. In this time the deviation caused by the earth's gravity would amount to about half a kilometer, whereas the meteor would have travelled 120 km to 720 km, depending on its velocity. Even in the extreme case of a meteor path tangential to the earth's surface, this deviation is less than can be detected by our present radar techniques. Other perturbing factors may be present, for example, explosions or bursts along the meteor path, but Whipple's photographic observations[2] show that these cause deviations from linearity rarely as great as a third of a degree. On the other hand, the assumption of a constant velocity was justified only because small decelerations could not be measured to any degree of precision with the instrumental techniques available at the time. Whipple[3] states

[1] D. W. R. McKinley and P. M. Millman, Can. J. Research A27, 53 (1949).

[2] Fred L. Whipple, Proc. Soc. Phil. Am. 91, 189 (1947).

[3] Fred L. Whipple, Revs. Modern Phys. 15, 243 (1943).

that the photographic meteors often do show decelerations of the order of 1 km sec^{-2}. The existence of similar decelerations in the moving radar echoes has been suspected.

In the spring of 1949 a new timing system was put into operation to supply precision seconds pulses to all four stations in the Ottawa area network, (Ottawa, Arnprior, and Carleton Place radar stations, and South Gloucester Doppler receiving station). The old system used the seconds pulses from the transmissions of CHU[4] on 3.3 Mc/sec. There was an average timing error, or jitter, of 0.02 sec between consecutive seconds, due partly to the mechanical keying arrangement at CHU and partly to the destructive interference effects between the sky wave and the ground wave. Furthermore, the CHU seconds signals each consisted of a burst of 1000 cps tone, which was a fundamental limitation to short interval timing precision. The use of CHU transmissions was, therefore, discontinued, except as a stand-by service.

A Western Electric 100-kc/sec crystal, with an intrinsic accuracy of 1 part in 10^8, has been installed at the Ottawa radar station, and precision seconds pulses are produced, by means of electromechanical gating and counting-down circuits,[5] which have intersecond jitters of less than a microsecond. These seconds pulses are used to key a transmitter on 220 Mc/sec which emits one pulse each second, of 20-microseconds duration and 5-kw peak power, from an antenna on top of a 200-foot tower. Direct line-of-sight is obtainable to all the stations, and no skywave troubles are encountered. The timing pulses, as received at all stations, now have intersecond jitters of the order of 10^{-4} sec or smaller, as compared with the previous jitter of 2×10^{-2} sec. Thus, the intrinsic errors in the timing are a hundred times smaller than the smallest reading errors in the photographic radar records. The film has been speeded up from one to four inches per minute to improve the reading resolution. At present, the timing precision is limited primarily by the recurrence frequency of the pulses from the high power radar transmitter which, in effect, samples the echo 120 times a second.

In the measurement of range some improvements have been effected by expanding the range scale so that the over-all range coverage is 260 km instead of 360 km, and by calibrating the range sweeps simultaneously at all stations from the Ottawa station. A pulsed crystal oscillator is used to produce precision 20-km range pips at Ottawa, and, then, at intervals, these standard range pips are sent out from the 220-Mc/sec time signal transmitter, so that all stations can photograph the standard pips simultaneously with their own locally generated pips.

Since the time the above instrumental changes were made, the Ottawa radar records have been searched for meteor echoes with structure and definition that would be satisfactory for deceleration measurements, but very few have been found. This is partly because the radar system has been operated during the past year with reduced noise level on the display, which improves the echo definition but decreases the effective sensitivity to weak signals. The average operating time each month has been about 72 hours. As remarked in the previous paper,[1] there is, in any case, a limited number of echoes having a well-defined h characteristic capable of yielding a velocity determination with a probable error of the order of one or two percent. The meteor should be quite bright, by visual standards, and should come from a radiant low in the sky to ensure that a hundred-odd kilometers of its path will lie in the M region.[6] Less than one in ten of such echoes that are acceptable for velocity measurement is suitable for deceleration measurement, as the precision required in the reduction work is an order of magnitude higher.

II. ANALYSIS OF RADAR RECORDS

We shall assume, as before, that the observed path of the meteor is a straight line in space. Allowance could be made for the acceleration due to gravity, 9.8 m sec^{-2}, but this effect is so small over the observed path in comparison with the expected decelerations that the correction can be neglected. On the other hand, we shall no longer assume a constant meteor velocity. Suppose the t_0 time is known from the radar record; let $\tau=t-t_0$ be the relative time along the path measured from t_0. The distance along the path, measured from the t_0 point, is $s=s(\tau)=\pm(R^2-R_0^2)^{\frac{1}{2}}$, the negative sign being used prior to t_0 and the positive sign after t_0. The ranges are read from the photograph at intervals of one-tenth or one-fifth of a second, and the best values for R_0 and t_0 are deduced from this data, assuming only that the points should be on a smooth curve, deviating slightly from a true hyperbola. A graph of s against t may show a slight curvature, if deceleration is present. A mean value of the velocity, v_m, can be found from the slope of the best line through the points of this graph. This technique is equivalent to the $R\dot{R}$ method of measuring velocity described in the previous paper.[1]

From the computed values of s the first differences may be obtained, and a plot of $\Delta s/\Delta t$ will show the variation of the instantaneous velocity along the path of the meteor. Second differences might, then, be used to show any change in deceleration, but so far the available observational data have usually not been accurate enough to reveal any significant changes in deceleration. Therefore, one might assume as a second order of approximation that the deceleration is constant over the observed path, and that $s=s(\tau)=v_0\tau+\frac{1}{2}a\tau^2$, where v_0 is the velocity at t_0 and a is the deceleration (numerically negative). The slope of the straight line through the points of the graph of s/τ *versus* t is then one-half the deceleration, and the intercept at t_0 yields the value v_0.

[4] M. M. Thomson, J. Roy. Astron. Soc. Can. **42**, 105 (1948).
[5] D. W. R. McKinley, Can. J. Research **F27**, 49 (1949).
[6] D. W. R. McKinley and P. M. Millman, Proc. Inst. Radio Engrs. **37**, 364 (1949).

This method is of some practical value in cases where the points on the $\Delta s/\Delta t$ graph are scattered widely, though it tends to give a false impression of smoothness and may conceal changes in deceleration.

Some examples selected from the Ottawa radar records will be used later to illustrate these methods. However, as Doppler whistle records are also available for each of the selected echoes, it may be advisable to insert a few remarks concerning the Doppler technique. This outline will be sketchy as a more complete account will appear elsewhere.[7]

III. DOPPLER METHOD OF MEASURING METEOR VELOCITIES

In contrast to the radar method of observing meteors where short bursts of energy are emitted and the time of arrival i.e., the range, of the echo is measured, the Doppler transmitter emits radiation, either in rapid pulses[8] or continuously,[9] and the receiver records the instantaneous echo amplitude rather than the echo range. In practice, the Doppler transmitter and receiver should be separated by several miles to attenuate the strong direct ground signal, if continuous-wave radiation is used, though this is unnecessary with a pulse transmitter. If one assumes the moving meteor to be creating a column of ionization that acts as a line scattering source, it may be shown[10] that the instantaneous intensity at the receiver of the signal from the meteor, in the absence of a ground wave, is

$$I_r = F_r^2(C^2+S^2), \tag{2}$$

where F_r is a constant proportional to the ultimate field strength of the echo and C and S are the conventional Fresnel integrals of optical theory, with limits of integration along the path taken effectively from minus infinity to the value of s at the selected instant. This is identical to the optical diffraction pattern at a straight edge, and by plotting Eq. (2) as intensity *versus* time, it can be seen that oscillations of decreasing amplitude and increasing frequency will occur *after* the t_0 point. The term "Doppler whistle" is applied to these oscillations, because if the output of the receiver is applied to a loudspeaker, an audiofrequency whistle of varying pitch is heard.

If a direct ground wave is present,[11] then the resultant intensity is,

$$I_{r+d} = (F_r C + F_d \cos\psi)^2 + (F_r S - F_d \sin\psi)^2, \tag{3}$$

where F_d represents the ground wave field strength, and ψ is the phase angle between the ground wave and the wave returned from the meteor at time t_0. A plot of I_{r+d} against time shows a Doppler whistle of increasing amplitude and decreasing frequency on the approach side of t_0 as well as the increasing pitch whistle on the receding side. In practice, the presence of the ground wave is beneficial as the approach whistles are usually clean and easy to measure, while the receding whistles, if present at all, are frequently distorted by a low frequency, large amplitude variation in echo strength that is associated with some slow relative motion of the ionized column as a whole, after the meteor itself has vanished.

By using Cauchy's approximations for Fresnel's integrals, it can be shown that the oscillatory part of the intensity, Eq. (3), has the form $(K/s)\cdot\sin[(2\pi s^2/R_0\lambda)+\psi]$, over a limited region that excludes t_0 and at the same time is not too remote from t_0 to render invalid the original approximations employed in setting up Eqs. (2) and (3). Nearly all of the measurable Doppler whistle records lie well within this region. If we assume the velocity to be constant over the short duration of the whistle record, i.e., $s=v\tau$, then the instantaneous frequency of the oscillatory term is given by,

$$f = 2v^2\tau/R_0\lambda, \tag{4}$$

and the slope of a graph of f against τ is proportional to the square of the velocity. Manning[12] deduced the same expression as given by Eq. (4), though from a different physical hypothesis. Manning's theory breaks down in the neighborhood of the t_0 point, whereas the diffraction theory, which was first suggested by Herlofson and elucidated by Lovell and Clegg[10] for the case of zero ground wave, appears to fit the observations better. The linear relationship between f and τ holds to a fraction of one percent, if the ten or twelve oscillations immediately adjoining the t_0 point are excluded. This was done in the examples to be discussed here, where up to several hundred oscillations could be read. On the other hand, if the total number of recorded oscillations does not exceed ten or fifteen, as in the majority of cases, the diffraction theory must be applied more rigorously to determine the velocity accurately, although a rapid measurement with an accuracy of a few percent may be made by using a slightly different approximate formula.

The radar value of R_0 is used in Eq. (4), but owing to the 7.5-km separation of the Doppler transmitter located at Ottawa and the receiver located at South Gloucester, the radar R_0 will differ slightly from the true Doppler R_0. The sign and magnitude of the correction to be applied to the radar R_0 depend on the position of the meteor in the sky, and, in the worst cases, it amounts to about ± 1 percent in the measured velocity. If the meteor position is known, the correction can easily be applied, otherwise the probable error of the whistle

[7] D. W. R. McKinley, Astron. J. **54**, 179 (1949). (This is an abstract of a paper presented at the Eighty-first Meeting of the Am. Astron. Soc.; Ottawa, June 1949. The full paper, with additional data, will appear in the Astrophysical Journal.)

[8] J. G. Davies and C. D. Ellyett, Phil. Mag. Ser. 7, **40**, 614 (1949).

[9] Manning, Villard, Jr., and Peterson, J. Appl. Phys. **20**, 475 (1949).

[10] A. C. B. Lovell and J. A. Clegg, Proc. Phys. Soc. **60**, 491 (1948).

[11] The extension of the diffraction theory to include the case of the ground wave is due to E. L. R. Webb, NRC, Ottawa.

[12] L. A. Manning, J. Appl. Phys. **19**, 689 (1948).

measurement must include the maximum value of the range correction.

The Doppler transmitter has a continuous wave output of about 1.5 kw on 30.02 Mc/sec, feeding a half-wave horizontal dipole mounted a quarter-wave above ground. A similar dipole is used at the receiving site, the axes of the two dipoles being nearly collinear to reduce the direct ground wave. The output of the receiver is applied to a cathode-ray tube as a vertical amplitude modulation of a horizontally moving trace, and is photographed by a camera which moves the film in a vertical plane continuously past the lens at 4 inches per minute. This film speed is the same as the radar film, so that comparison of the two records is facilitated. The cathode-ray tube trace moves from the bottom of Fig. 1(c) to the top in 0.25 sec, flashes back in about 0.001 sec and repeats, while moving slowly to the right. The same seconds pulses that appear on the radar record, Figs. 1(a) and 1(b) also appear on the Doppler trace, but while the time resolution of the radar record is of the order of 0.01 sec that of the Doppler is 0.001 sec. Frequencies up to 800 cps can be measured. Standard 100-cps calibration pips are applied to the trace from time to time, because the sweep is not strictly linear, and a detailed calibration of the sweep speed must be made before reducing the meteor whistle. A second type of recording technique is available, wherein the amplitude modulation is applied horizontally to a stationary spot on a cathode-ray tube and the film is drawn vertically past the lens at 4 inches per second by a synchronous motor. Apart from encountering much less difficulty in calibrating the trace, the frequency resolution is increased to the order of 1500 cps. However, this system is used only on special occasions and was not operating during the periods in which the illustrative meteor echoes occurred.

IV. MEASUREMENT OF VELOCITY AND DECELERATION

The three examples of meteor echoes shown in Figs. 1, 4, and 5 were selected because the clarity and duration of the moving part of the radar echo (*h* characteristic), permitted fairly accurate range and time measurements to be made. The Doppler records have been included for the additional information they provide, although these particular records are by no means striking. In fact, all the radar and Doppler photographs shown here suffer from interfering signals, instrumental defects, loss of detail in reproduction, and even film scratches. The reduction steps and the associated graphs are given in some detail for *Meteor I* but, apart from some relevant remarks in each case, the data for *Meteor II* and *Meteor III* are merely tabulated in the summary, Table I.

Meteor I

This meteor occurred at 09 hr 06 min 18 sec EST on September 20, 1949 and appears to have been a sporadic

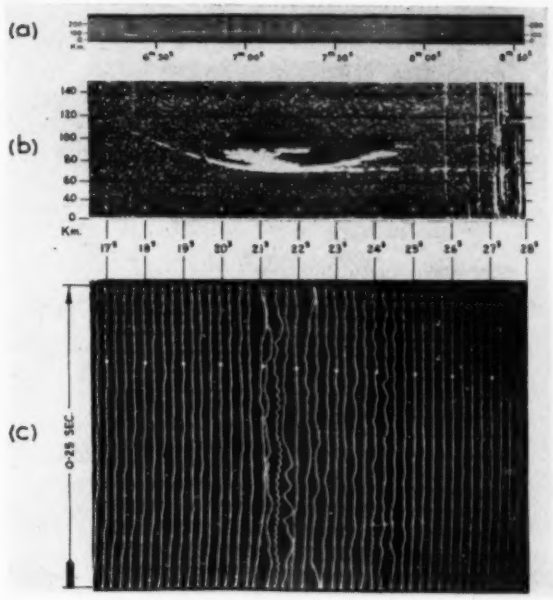


FIG. 1. *Meteor I*. 09 hr 06 min 18 sec EST, September 20, 1949. (a) Ottawa Radar Record, 6 min to 8 min 30 sec showing *Uhe4* echo, over-all duration 154 sec. (b) Ottawa Radar Record, detail of *Uh* part of echo. (c) Ottawa—South Gloucester Doppler Record, showing high frequency oscillations just prior to t_0 point.

meteor, with a radar echo of type *Uhe4*[13] [Figs. 1(a) and 1(b) and Table I]. The time duration of its *h* characteristic from 18 sec to 28 sec is the longest that we have observed [Fig. 1(b)]. This represents a radar path length of about 190 km, and, although longer path lengths have been measured,[1,14] the corresponding velocities were greater. The *h* characteristic fades at 24.4 sec but reappears as a faint line between 25.8 sec and 28.0 sec. The vertical streaks are caused by the transmissions of some unknown distant station, probably reflected from the ionosphere in this case, though the remote signals are often reflected from the local ionized cloud that expands in the wake of the meteor. The over-all duration of the *e* characteristic [Fig. 1(a)], is 154 seconds, and the most enduring components lie at ranges between 107 and 123 km. They do not appear strongly until about 10 to 15 seconds after the passage of the meteor.

Figure 1(c) shows the Doppler record for *Meteor I*. The whistle oscillations were first resolved at 20.85 sec. They may exist earlier, but the amplitude would be too small and the frequency too high to be detected. The large amplitude, low frequency oscillations superposed on the high frequency whistle have no connection with the velocity of the meteor but are probably caused by movement of the ionized column already formed, whether due to winds, to radial expansion of the column, or to some other agency.

[13] The nomenclature and classification of meteor radar echoes are given in reference 6.

[14] P. M. Millman and D. W. R. McKinley, J. Roy. Astron. Soc. Can. 42, 121 (1948).

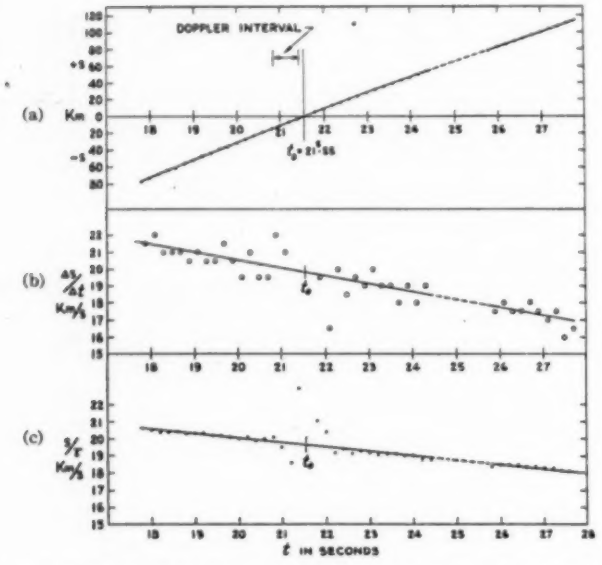


FIG. 2. *Meteor I*, Radar graphical data. (a) s plotted against t. (b) $\Delta s/\Delta t$ plotted against t. (c) s/τ plotted against t.

Figure 2(a) shows the *s versus t* plot obtained from the radar record. This has a perceptible curvature, but if the quality of the observational data had been poorer, the best that could have been done would be to draw a straight line through the points. This would yield a mean velocity v_m. The mean velocity thus obtained in this case is 19.0 km/sec. The time interval of the Doppler whistle is shown on Fig. 2(a) for comparison purposes. The Doppler t_0 time is 21.562 sec, compared with the radar t_0 time, 21.55 sec. From these figures, it is deduced that the meteor has probably traveled nearly parallel to the right bisector of the line joining the Ottawa and South Gloucester stations. Had the meteor been travelling nearly parallel to the Ottawa–South Gloucester axis the difference in t_0 times could have been as much as ± 0.18 sec.

Figure 2(b) shows the plot of $\Delta s/\Delta t$. The points are scattered, especially near t_0 where the range increments are small, but a general trend is apparent. The velocity at the beginning of the radar path can be read from this graph, $v_b = 21.5$ km/sec, and that at the end, $v_e = 17.0$ km/sec. Some curvature may be present over the path, and there is a definite suggestion of a more sudden deceleration at the end, but the coarseness of the data does not appear to warrant any treatment other than drawing a straight line through the points. The mean deceleration of the meteor over the path is given by the slope of this line, $a = -0.47$ km sec^{-2}.

In Fig. 2(c) the same data is replotted as s/τ *versus* t. Again, the relationship appears to be quite linear, the slope being -0.24, which is half the meteor deceleration.

The Doppler whistle measurements are plotted in Fig. 3 as frequency against time. Each of the circled points represents ten oscillations, each set having been measured in two groups of five, while the crosses represent four oscillations each, measured a pair at a time. Some parts of the trace were unuseable, but readable portions can be found on the original film as far as the three-hundredth oscillation prior to the t_0 point, with the aid of a low power microscope. The mean velocity over the measured time interval is $v(D) = 20.1$ km/sec. No appreciable curvature is apparent in the graph, Fig. 3, but this is not surprising, even though the slope of the line is proportional to the square of the velocity, because the total change in velocity over this short time would be only 0.3 km/sec, using the radar deceleration values. *Meteor I* was not observed visually, nor were radar records of it available from the other two radar stations for triangulation, so that its position in the sky is uncertain. However, the low value of R_0 suggests that it must have passed nearly overhead, hence, the error introduced by using the radar R_0 figure in the Doppler formula is unlikely to exceed ± 0.2 km/sec in the velocity. The measurement error in the Doppler whistle velocity reduction is also about ± 0.2 km/sec. Thus, the Doppler velocity is $v(D) = 20.1 \pm 0.4$ km/sec, which compares favorably with the radar velocity, $v_D = 20.0 \pm 0.2$ km/sec, at the mid-point of the Doppler interval.

From the radar record alone, some conjectures may be made concerning the heights along the meteor path. It is not unreasonable to associate the long-enduring echoes with the M region,[6,14] which extends in height from about 75 km to 115 km, with a maximum at about 95 km. The long-enduring, or e characteristic, parts of the echo extend in range from 107 km to 123 km, and definitely have their origin on the approach part of the meteor path. By extrapolating the observed path determined from the h characteristic, it is found that the range of 115 km, corresponding to the mean range of the observed e characteristic, occurs at time $t = 17.25$ sec. Let us assume tentatively that the height at 17.25 sec is in the lower part of the M region, somewhere between 75 and 95 km. At $t_0 = 21.55$ sec the range is 70.9 km, and this would be the maximum possible height at this instant. The actual height at time t_0 is likely to be a few kilometers less than 70.9 km, say between 65 and 70 km.

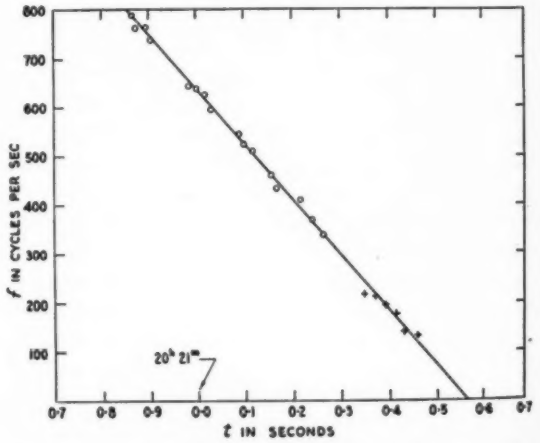


FIG. 3. *Meteor I*, Doppler graphical data. f plotted against t.

Using this data to draw a height-time graph, we find the spread in possible heights at the end of the path ($t=28.0$ sec), to be from 30 to 65 km, with the most likely terminal height being about 50 km.

Meteor II

This meteor occurred at 18 hr 39 min 55 sec EST, on April 21, 1949, as a type *Uhb* echo; see Fig. 4(a) and Table I.

There is a reasonable possibility that it is a Lyrid meteor, first, because of its velocity which varied from 49 to 43 km/sec, and second, because it appeared when the Lyrid radiant was just below the horizon, a condition that seems to favor the production of striking echoes of this type.[1,14] The corrected elevation of the mean Lyrid radiant at this time was $-3°$, but the cosmic spread of the radiant might be sufficient to allow this meteor to be a member of the Lyrid shower. However, it was not observed visually, nor was it triangulated by radar so the identification is uncertain.

The quality of the measurements made on the radar record, Fig. 4(a), was inferior to those made on *Meteor I*, so that the probable errors attached to the computed velocities and to the deceleration value, $a=-1.1\pm0.4$ km sec^{-2}, are larger (Table I). At the time this photograph was made, a defect in the driving mechanism caused the film to move jerkily, as shown by the uneven background shading, and an attempt was made to correct the observed time measurements accordingly.

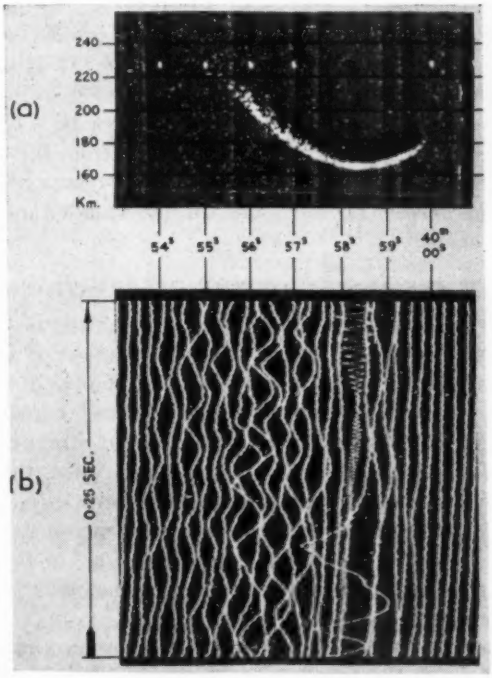


FIG. 4. *Meteor II*, 18 hr 39 min 55 sec EST, April 21, 1949. (a) Ottawa Radar Record, *Uhb* echo. (b) Ottawa—South Gloucester Doppler Record, showing high frequency oscillations, with envelope variations, just prior to t_0. Slow oscillations are also present for four seconds prior to t_0.

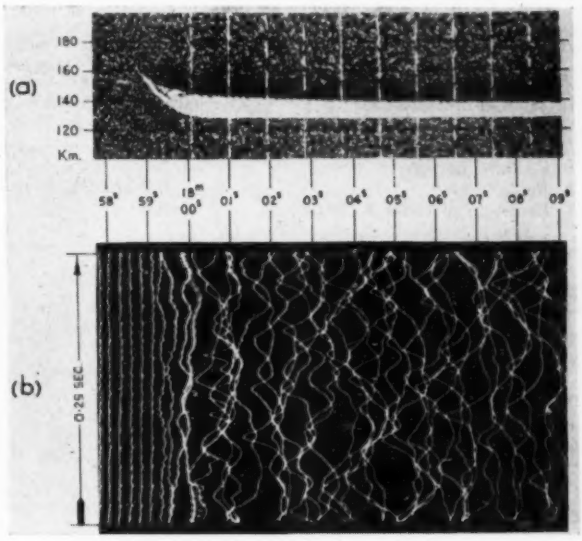


FIG. 5. *Meteor III*, 20 hr 17 min 58.7 sec EST, August 11, 1949. (a) Ottawa Radar Record, *Ahbf* echo, overall duration more than 85 sec. (b) Ottawa—South Gloucester Doppler Record, showing high frequency oscillations just prior to t_0. Slow oscillations of large amplitude commence 1 sec before the t_0 point and continue for more than 85 sec.

The Doppler whistle, Fig. 4(b), yielded a velocity, $v(D)=45.4\pm0.3$ km/sec, using the radar $R_0=164.4$ km. If it is assumed to be a Lyrid meteor and to have occurred at the general height level of 100 km, then the meteor can be located in either of two positions in the sky and the adjustment to the velocity to include the range correction will be either plus or minus 0.4 km/sec; hence, two velocity values are quoted in Table I. If it is not a Lyrid, then the maximum correction of ±0.45 km/sec must be included in the over-all probable error. The value, $v(D)=45.0$ km/sec, does happen to agree with the radar velocity, $v_D=45.1$ km/sec, at the midpoint of the Doppler whistle interval.

It is not clear why this meteor, which has produced such a striking *Uhb* echo, did not leave an enduring cloud of ionization as did *Meteor I* and *Meteor III*. Such echoes are not rare, but they are in the minority. A reasonable explanation might be that the observed path of the meteor was well above or below the center of the *M* region—in this case, it would likely be above, since visual and photographic evidence indicates that the faster meteors occur at higher levels.

Meteor III

This meteor, very likely a Perseid, occurred at 20 hr 17 min 58.7 sec EST on August 11, 1949, and produced a radar echo of type *Ahbf* [Fig. 5(a)]. All three stations in the radar network were operating at the time, as well as the Doppler system, and, although the probable errors of the path position elements were rather large because the Carleton Place record showed little or no *b* characteristic, the computed path lay within a few degrees of the Perseid radiant. The velocity range,

TABLE I. Summary of radar and Doppler data for three meteors.

	Meteor I	Meteor II	Meteor III
1. Time of appearance	09 hr 06 min 18 sec EST September 20, 1949	18 hr 39 min 55 sec EST April 21, 1949	20 hr 17 min 58.7 sec EST August 11, 1949
2. Identification	Sporadic	Possible Lyrid	Perseid
3. Radar echo type	*Uhe4*	*Uhb*	*Ahbf*
4. R_0, t_0	70.9 km, 21.55 sec	164.4 km, 58.42 sec	129.4 km, 00.28 sec
5. Duration of radar *h* characteristic	10.0 sec	4.4 sec	1.6 sec
6. Duration of measurable Doppler whistle	0.60 sec	0.27 sec	0.16 sec
7. Duration of radar *e* or *f* characteristics	154 sec	none	over 85 sec
8. Observed radar path length	190 km	200 km	90 km
9. Height above sea level			
at beginning of radar path	75–100 km	—	121 km
at end of radar path	30–65 km	—	91 km
10. Radar velocity, km/sec			
v_m—mean value	19.0	46.2	59.8
v_b—at beginning of path	21.5	49.0	61.0
v_e—at end of path	17.0	43.4	58.5
v_D—at mid-point of Doppler interval	20.0±0.2	45.1±0.6	58.7±0.7
11. Doppler velocity, $v(D)$, km/sec	20.1±0.4	45.0±0.3 or 45.8±0.3	58.3±0.4
12. Mean radar deceleration, a, km sec^{-2}	−0.48±0.05	−1.1±0.4	−1.5±0.3

61.0−58.8 km/sec (Table I), is characteristic of the Perseid stream. A third and much less tangible reason for assuming *Meteor III* to be a Perseid is that the radar echo "looks like a Perseid." That is, the *Abhf* echo with a broad *b* band seems in a qualitative way to be more characteristic of the echoes obtained during the Perseid shower than, say, during the equally prolific Geminid shower. From the quantitative evidence of path position and velocity, it seems safe to identify this meteor as a Perseid, without the intuitive aid of the echo form.

The receding portion of the ionized path of the meteor, if it exists, is completely masked by the strong *bf* characteristic. However, the *Ah* part of the echo is sufficiently well-defined to yield reasonably accurate values of velocity and deceleration (Table I). Triangulation of the lowest part of the enduring echo places this point on the path (approximately the R_0, t_0 point at Ottawa), at a height of 91 km above sea level and on a true bearing from Ottawa of 265°. By combining the known elevation of the Perseid radiant, 18° at this time, with the radar figures for the height at the Ottawa t_0 point and the velocity along the path, the height at the beginning of the observed path is found to be 121 km.

The vertical streaks on the record, Fig. 5(a), at intervals of approximately 0.9 sec are due to reflections from the meteoric cloud of the Arnprior radar transmitter, which appears to have been operating undesirably close to the Ottawa frequency at the time. However, because the pulse recurrence frequencies of the Ottawa and Arnprior transmitters were slightly different, the Arnprior echo moved rapidly across the Ottawa display without confusing the picture unduly. The over-all duration of the *f* characteristic was in excess of 85 sec; a second *Ahbf* echo, appearing at the same range 85 sec later, obscured the latter portion of the echo from *Meteor III*.

The Doppler whistle record, Fig. 5(b), is a good specimen for measurement purposes, though the reproduction unfortunately does not show the detail of the high frequency oscillations very well. The low frequency oscillations, or "body Doppler,"[9] that continue for the duration of the radar echo are very marked in this example. The range correction can be applied without ambiguity since the position of *Meteor III* is known. The velocity is $v(D)=58.3\pm0.4$ km/sec.

The mean deceleration for this meteor is -1.5 ± 0.3 km sec^{-2}. Despite its shorter time duration, the clarity of the *h* characteristic of *Meteor III*, compared with that of *Meteor II*, accounts for the smaller measurement error.

V. DISCUSSION OF THE OBSERVATIONS

The reality of the apparent decelerations of the meteors listed in Table I may be questioned at this stage. In this connection, it might be helpful to outline briefly some of our current views on radar and radio echoes from meteoric ionization. There are three general types of echoes received from meteors; first, the echo received from the very dense line source of electrons formed in the immediate wake of the meteor; second, the echo received from the expanding cloud of electrons some time later; third, the moving echo associated with the bright meteors.

The first echo type is the most common and is obtained from the small meteors observed at the lower radiofrequencies and from practically all the meteors observed at higher frequencies. Satisfactory agreement with experimental data has been obtained by applying diffraction theory to the radiofrequency energy scat-

tered from the dense column of free electrons being formed by kinetic collision processes immediately behind the meteor. The diameter of the column is assumed to be small compared with the wavelength. As the diffraction formula shows, Eq. (2), the intensity of the echo reaches a maximum just after t_0. The radar set and the Doppler system without a ground wave, both observe this increase in intensity, though the Doppler equipment resolves the fine structure of high frequency oscillations superposed on the average increase in strength. The situation is not fundamentally changed when the ground wave is added to the Doppler system, Eq. (3), although the record may appear to have been radically altered by the introduction of an additional fine structure of oscillations prior to t_0, and the average signal level after t_0 may either increase or decrease from the steady state level set by the ground wave, depending on the relative phase of the two waves arriving at the receiver. The basic echo strength formula developed by Lovell and Clegg[10] [see also Eq. (10)] from earlier work by Blackett and Lovell[15] involves the line density of the free electrons and does not require that any time should elapse to allow the column to expand. Lovell has shown that the formula holds over a wide range of wavelengths, and the work of Davies and Ellyett,[8] as well as our own records, show clearly that the variations of the echo amplitude agree closely with that predicted from simple diffraction theory.

On the other hand, the second type of echo takes the form of quasi-stationary echoes of comparatively long durations (*e* and *f* characteristics), which may be delayed after the passage of the meteor by intervals ranging from small fractions of a second to as much as fifteen or twenty seconds. A phenomenological theory to account for these echoes has been proposed in an earlier paper,[6] in which it was suggested that some localized parts of the *M* region are more conducive than others to the production and maintenance of the meteoric ionization. Some time after the passage of the meteor, these dense regions on the meteor path expand to clouds many wavelengths in diameter, thus forming targets of detectable size. The delay in the appearance of the echo was shown from geometrical considerations, and confirmed by observational evidence, to be greatest for those parts along the path most remote from the t_0 point on either side. It was further suggested that the shortest delays would occur for echoing regions located near the upper levels of the *M* region, where diffusion of the electrons might be expected to be most rapid. The reflection mechanism in the case of the *e* and *f* characteristics may be presumed to be that of total reflection from a relatively large volume of ionized gas, the electron density being above the critical density for the wavelength employed. Alternatively, one could assume the echo to be the summation of waves scattered from a large number of discrete scattering centers spread throughout this volume.

The third type is a moving echo (*h* characteristic), with a velocity apparently equal to the actual velocity of the meteor and which, in general, is associated with the brighter meteors. The *h*-characteristic echo may be obtained from meteors moving directly toward or almost directly away from the observer, that is, from parts of the path remote from the t_0 point. It is unlikely that this echo can be explained entirely by diffraction from the dense line of ionization produced by collision. The intensity of diffracted echo varies inversely as the distance from t_0, and for remote points it will be very weak and will increase steadily as t_0 is reached, whereas the *h*-characteristic echo is often fairly strong and remains constant or even diminishes in intensity as t_0 is approached. Furthermore, if a diffracted echo were received at all from a given line segment of the path remote from t_0, it should continue to be seen as the line source expands. A feature of the majority of the *h*-characteristic echoes is that the echo disappears very quickly, usually within the resolution limit set by the transmitter pulse width, and some time later echoes of the second type appear with delays ranging up to several seconds, although the delays for parts of the *b*-characteristic echoes may be vanishingly small. It is improbable that the echo could be associated with a sonic shock wave, because the cone angle of the shock wave would be much less than a degree, and it is difficult to see how energy could be scattered back to the observer from such a shape, whether approaching or receding, even if the dielectric discontinuity introduced by the sound wave boundary were adequate for the reflection or scattering of the radio wave.

The most attractive hypothesis yet advanced to explain the *h*-characteristic echo appears to be that ultraviolet light from the head of the meteor can instantly ionize the air at a considerable distance from the meteor.[6] The density of the ionization would be small compared with that of the column formed by meteor-air collisions but, when observed on the lower radio-frequencies, it could momentarily be above the critical density throughout a sizeable region around the meteor head. The echoing area of this tenuous head cloud should be relatively independent of the angle at which it is viewed by the observer, whether approaching or receding. Furthermore, because the cloud is presumed to be created full-blown, as it were, by light radiation, and not by diffusion of the dense line ionization, no delays should be involved in its detection, regardless of the height of the meteor. The echo is obtained from this head cloud by total reflection processes, until the density drops below the critical which usually occurs very quickly. Some time later, measured in milliseconds to seconds, the *b*-, *e*-, and *f*-characteristic echoes may be obtained from the expanding parts of the dense core. If the point of reflection is at or near the t_0 point, then the dense column will be seen instantaneously by virtue of

[15] P. M. S. Blackett and A. C. B. Lovell, Proc. Roy. Soc. (London) **A177**, 183 (1941).

the diffraction effect (first echo type). As the t_0 point is approached by a bright meteor, the diffraction echo increases in amplitude, and in the t_0 region it may greatly exceed the relatively constant amplitude echo from the moving head cloud.

If a Doppler system with ground wave is used to detect the head cloud echo at points relatively remote from t_0, a Doppler whistle should be obtained by simple *interference* effects between the moving echo and the ground wave, rather than by *diffraction* effects from a moving line source combined with interference from the ground wave, Eq. (3). Manning[12] has worked out the theory for the moving cloud case, though he applied it to the whistle obtained from the average meteor near t_0, which we now believe to be explained more satisfactorily by the diffraction theory. Manning's theory would apply more accurately to the rather rare cases of the moving head cloud or h-characteristic echo. Just before t_0 the Doppler system with ground wave will detect high frequency oscillations, both by diffraction from the moving line source and by interference from the head cloud. The frequencies will be almost identical and the amplitudes may be comparable, but the relative phases will change slowly, because the reflecting surface of the cloud is somewhat indeterminate; hence, the resultant whistle record will represent the sum of the two oscillations, reinforcing in some places and cancelling in others. This can be seen in Figs. 1(c), 4(b), and 5(b), though it is most apparent in Fig. 4(b), where the envelope of the oscillations has varied through several maxima and minima. It is true that the same effect might be due to diffraction echoes from two particles of equal size moving near each other on parallel paths with almost the same velocity. However, the "modulation envelope" effect occurs quite frequently with this particular type of echo having a pronounced radar h characteristic, and, in general, we prefer the explanation given above involving one particle and two echoing mechanisms.

The close agreement between the velocities measured at the same point on the meteor path by the radar range-time and the Doppler amplitude-time methods is apparent from a comparison of v_D and $v(D)$, Table I. It is obvious that the Doppler system in its present stage of refinement can sample the velocity only over a short path interval. While this may be a drawback for the purpose of measuring decelerations, the short path requirement does mean that many more meteor velocities can be measured by amplitude-time methods (including both continuous wave and pulse systems), than can be determined by range-time methods. However, in measuring a large number of velocities of shower meteors, all of which presumably have nearly the same velocity outside the earth's atmosphere, a spread in velocities will be found which, in addition to the usual measurement errors, will include a real variation in the observed velocities.[7] Those shower meteors observed with t_0 points near the top of the M region will yield measured velocities somewhat higher than those with t_0 points in the lower levels of the M region, because the latter will suffer more deceleration on the average.

Assuming that, on the basis of the foregoing arguments, the reality of the observed decelerations has been accepted, then questions arise concerning the apparent constancy and the order of magnitude of the effect. One would expect that the deceleration should increase as the meteor travels downwards into the denser air. Photographic evidence shows that small increases in deceleration do occur along the paths, and rapid increases in deceleration are sometimes observed at the end of the paths, just before the meteors are completely burned up. The inclinations of the meteors listed in Table I are relatively low, hence, any increase in deceleration with decrease in height might be too small to be observed with the present technique. It is quite likely, too, that the radar has not seen the actual end of the meteor paths in these cases. Our combined observations[6] indicate that the visual and photographic paths are, in general, longer than the radar paths.

An attempt has been made to rework the velocity measurements made on about twenty radar echo specimens obtained prior to the spring of 1949, but the quality of the records proved to be inadequate. Only rough orders of magnitude could be obtained which suggested that the decelerations ranged from zero to several km sec^{-2}.

The technique for measuring meteor decelerations outlined here has a limited application as only the very bright meteors travelling at low inclinations can produce suitable records. The information obtained under these special conditions may not apply to fainter meteors coming from radiants higher in the sky. A promising attack on the latter problem might be to use two Doppler systems, separated by several kilometers, so that the velocity of the meteor could be measured accurately at two points on its path. The instrumental precision and resolution of a single Doppler system might be improved to the extent that a change in velocity could be detected over the path of the whistle. As the discussion of the next section will indicate, the rate of change of the deceleration of the meteor would be a very useful quantity to have available, but even greater refinements in the radar and Doppler techniques will have to be made before variations in the decelerations can be determined with sufficient accuracy.

VI. APPLICATIONS—GRADIENT OF ATMOSPHERIC DENSITY

In computing the atmospheric density from photographic meteor observations, two basic equations have been used which appear to be reasonably well established. Without going into the history in detail, we shall summarize these from Whipple's[3] review paper as follows:

$$a = -k_1 \rho m^{-\frac{1}{3}} v^2, \tag{5}$$

$$dm/dt = -k_2 \rho m^{\frac{2}{3}} v^3, \tag{6}$$

where m is the mass of the meteoric particle, and ρ is the atmospheric density. In Eq. (5) the constant, k_1, is a function of the shape and density of the particle. Whipple uses the value $k_1=0.5$, but suggests that it may be as great as 0.9. In Eq. (6) the constant k_2 is a function not only of the shape and density but also of the energy required to melt and vaporize 1 gram of the material of the particle. Whipple adopts a value, $k_2=2\times10^{-12}$ cgs units as representative of an average meteor, though there may be a wide individual variation.

By differentiating Eq. (5) and using Eq. (6), one can eliminate m and dm/dt to obtain an expression for the logarithmic gradient of the atmospheric density as follows,

$$\frac{1}{\rho}\cdot\frac{d\rho}{dH}=-\left(\frac{k_2}{3k_1}-\frac{2}{v^2}+\frac{1}{a^2v}\cdot\frac{da}{dt}\right)a\ \mathrm{cosec}h, \qquad (7)$$

by assuming $v\cdot dt=ds=-dH\cdot\mathrm{cosec}h$, where dH is an element of height, and h is the elevation of the meteor radiant. If $da/dt=0$, we have the anomalous result that the gradient is positive (a is negative), except for very slow meteors. Unless the basic reasoning underlying Eqs. (5) and (6) is in error, which is unlikely, we are forced to conclude that the radar echo measurements should show negative rates of change of deceleration,[16] though these may have been obscured by the observational errors. *Meteor I* has the best data for a, and, while the radiant elevation was unknown for this meteor, we can deduce from Eq. (7), using the meteor elements listed in Table I, that da/dt should be at least -0.05 km sec^{-3} to make the gradient negative. The plot of $\Delta s/\Delta t$, Fig. 2(b), for this meteor does show evidence of a change of slope of this magnitude, though it would be difficult to assess it quantitatively. At this stage in the experimental development, it does not appear feasible to attempt to compute the atmospheric gradient from measurements of v, a, and da/dt, as the percentage error in da/dt should be of the same order as the errors in v and a to obtain useful results.

It may be noted that the absolute temperature is inversely proportional to the logarithmic density gradient. A technique for measuring da/dt accurately would be highly desirable, since these two important physical quantities could then be determined solely from a knowledge of the geometry of the meteor path.

VII. APPLICATIONS—ECHO STRENGTH AND IONIZING EFFICIENCY

Another approach to the deceleration problem appears to be more productive at present. In the photographic work on meteors it is assumed that the luminosity is proportional to the kinetic energy of the mass lost per second multiplied by a luminous efficiency factor, which is taken to be a linear function of the velocity. An analogous relationship in the radio work would be to assume that N, the number of electrons produced per second, is proportional to the rate of loss of kinetic energy of the meteor multiplied by an ionizing coefficient k_3, which may be a function both of the velocity and of the height. Suppose that ionization potentials of 15 electron volts are representative or that 2×10^{-11} erg are required to produce each electron on the average. We then have,

$$N=-\frac{1}{2\times10^{-11}}\cdot k_3\cdot\frac{1}{2}\cdot\frac{dm}{dt}\cdot v^2. \qquad (8)$$

Substituting for dm/dt from Eq. (6) and, at the same time, eliminating m by using Eq. (5), we have

$$N=1.25\times10^{-2}\cdot k_3\cdot\rho\cdot a^{-2}\cdot v^9, \qquad (9)$$

using the previous values of k_1 and k_2.

N may be obtained from a knowledge of the radar parameters. Lovell[10] has established a formula for the number of electrons produced per cm of path, N/v;

$$\frac{N}{v}=5.5\times10^{13}\cdot\frac{1}{G}\cdot\left(\frac{\epsilon R_0^3}{P\lambda^3}\right)^{\frac{1}{2}}, \qquad (10)$$

where G is the antenna gain, ϵ the received echo power in watts, and P the transmitted peak power in watts. There have been occasions when the Ottawa radar system parameters have been measured with fair accuracy, and while this was not done specifically during the routine periods in which the meteors described here occurred, the operating conditions were not likely to have differed substantially from the calibrated conditions. The receivers were normally adjusted so that video clipping began at 12 db above the average unity signal-to-noise level of 2×10^{-14} watt and saturation of the receiver occurred at 24 db above noise. *Meteor III* is the only one that has a completely determined path, hence, the data for this meteor will be used. It is apparent from the photograph, Fig. 5(a), that at the R_0, t_0 point the echo strength has at least reached the saturation level of 5×10^{-12} watt and may have exceeded it, though not by a large factor. We shall, therefore, use this figure as a minimum value of ϵ in Eq. (10). The antennae were half-wave dipoles mounted a quarter-wave above ground, and, at the 45-degree elevation angle that *Meteor III* was viewed from Ottawa, the gain, G, was approximately unity. The other parameters had values as follows: $P=2\times10^5$ watts; $\lambda=9.1$ m; $R_0=129.4$ km. The computed minimum value for N is, then, of the order of 3×10^{18} electrons per second. From Whipple[3] we find $\log\rho=-7.9$, or $\rho=1.26\times10^{-8}$ g/cc, at 91 km (the height at the t_0 point where Lovell's formula is applicable). At t_0 the velocity of *Meteor III* was 58.5 km/sec. The corresponding deceleration at t_0

[16] For a very large meteor da/dt may be expected to become positive eventually, if the body survives long enough to approach a limiting velocity, but this effect may never be observed, first, because such large meteors are extremely rare, and, second, because the change in sign might be expected to occur at unusually low heights where there may be some doubt concerning the radar detectability of the ionized trail.

was probably greater than the mean measured value of -1.5 km sec^{-2}, but this will suffice as a minimum value. Substituting in Eq. (9), we find $k_3=3.4\times10^{-7}$. This is almost certainly an extreme lower limit for k_3, as all the probable variations in the variables of Eq. (9) tend to increase k_3. For example, it is likely that N might be increased by a factor of two or more, that a at the t_0 point might be two or three times the mean value, and that v would be slightly less (although from the nature of the reduction process, it is unlikely that v at t_0 would be less than 58.0 km/sec). Thus, k_3 may lie between 10^{-5} and 10^{-6}. This means that between 10^{-5} and 10^{-6} of the rate of loss of kinetic energy is converted to ionization for a meteor moving at 60 km/sec and at a height of 91 km.

A similar procedure can be followed for *Meteor I*, though with considerably more uncertainty, as we shall have to assume a height at t_0 between 65 and 70 km, say, with a corresponding density, $\rho=2\times10^{-7}$ g/cc. On the other hand, the observed value of the mean deceleration is more likely to be the value applicable at t_0. From the radar measurements, the computed minimum value of N for *Meteor I* is 6×10^{16} electrons per second, and the resultant k_3 is 3×10^{-9}. That is, k_3 is of the order of 10^{-8} for a meteor moving at 20 km/sec at a height of about 68 km. *Meteor II* cannot be used as its height is unknown.

The form of Lovell's formula, Eq. (10), has been reasonably well verified for a wide range of wavelengths and, also, for some variations in the other parameters. However, the absolute value of the constant, 5.5×10^{13}, which contains the scattering cross section of the free electrons, may be difficult to confirm experimentally. Or, if the value is correct, the possibility remains that only a fraction of the total number of electrons produced initially can be regarded as free electrons for scattering purposes. A further point to be considered in this discussion is that Lovell's formula applies to a long thin line source of free electrons. If the volume density of the electron cloud is above the critical density for the radiofrequency employed, Herlofson[17] suggests that plasma oscillations may occur and that the received echo may be expected to be much stronger than Eq. (10) would predict for a meteor producing a given number of electrons per second. The meteors described here were certainly brighter than average magnitude, and, hence, there is a possibility that Lovell's formula may not apply. However, if one adopts Herlofson's reasoning, the rate of production of electrons required to account for the observed signal strength should be considerably less, and the values of the ionizing coefficient would be correspondingly reduced. In this analysis, any contribution from the ionized head cloud created by ultraviolet radiation has been neglected, as observations indicate that the echo from the dense line source should be greater at the t_0 point.

[17] N. Herlofson, Observatory **68**, 226 (1948).

VIII. APPLICATIONS—LOSS OF MASS AND IONIZING EFFICIENCY

The order of magnitude of the mass of the meteoric particles may be found from Eq. (5), although its use in these cases is limited to points at or near the mid-point of the observed path where the mean deceleration figure applies, otherwise absurd results may follow. The ratio of the masses at any two points on the path can then be obtained by eliminating ρ from Eqs. (5) and (6) and integrating:

$$m_1/m_2=\exp[2\times10^{-12}(v_1^2-v_2^2)]. \tag{11}$$

Another expression for the rate of change of mass may be deduced from Eq. (6) by substituting for m from Eq. (5):

$$dm/dt=-5\times10^{-13}\rho^3a^{-2}v^7 \text{ g/sec}. \tag{12}$$

If we assume the particles to be spheres of stone with a density of 3.4, then the diameter, D, of the meteor is $0.83\ m^{\frac{1}{3}}$.

In particular, *Meteor III* had a velocity of 59.8 km/sec at the mid-point on its observed path, which was at a height of 106 km with a corresponding density, $\rho=2\times10^{-9}$. From Eq. (5) the mass was 1.4×10^{-2} g at this point, and from Eq. (11) the masses at the beginning and end of the observed path were 0.3 g and 10^{-3} g, respectively. From Eq. (12) the rate of loss of mass at the mid-point was 5×10^{-2} g/sec. The diameter of the particle at the mid-point of the path was 0.2 cm. Similarly, tentative values may be obtained from *Meteor I*, referring to the t_0 point where the air density is assumed to be 2×10^{-7}. The mass at t_0 was 500 g, the diameter 7 cm, and the rate of loss of mass 200 g/sec. The initial and final observed masses were 2000 and 60 g.

From the above data, an estimate may be made of the number of electrons produced per second. If we assume that the mean atomic weight of the meteoric particles was 30, then, in the case of *Meteor I*, the number of atoms released per second at the t_0 point was $200\times6.02\times10^{23}/30=4\times10^{24}$. The experiments of Sutton and Mouzon[18] suggest that the probability for the production of an electron by a meteor atom before it is stopped may be of the order of 10^{-2}. Hence, *Meteor I* should have been producing 4×10^{22} electrons per second at the t_0 point. While the observational data is admittedly very rough, this result is many orders of magnitude in excess of the figure, 6×10^{16} electrons per second, computed from Lovell's formula. The deceleration of *Meteor III* at the t_0 point is rather uncertain. Substituting $a=-1.5$ km/sec in Eq. (12) gives a rate of 10 g/sec, but the deceleration at t_0 is likely to be two or three times the mean value, hence, a fair estimate for dm/dt at t_0 might be of the order of 1 g/sec or less. This would correspond to the order of 10^{20} electrons per second, in somewhat closer agreement with the radar value, 3×10^{18}. Many factors have been omitted, particularly the velocity and height effects, and it does not

[18] R. M. Sutton and J. C. Mouzon, Phys. Rev. **37**, 379 (1929).

seem reasonable to expect that the laboratory results can be applied so loosely to meteoric ionization processes without further information. More experimental data are needed on the ionizing properties of 500- to 1000-electron-volt atoms moving in air of density from 10^{-6} to 10^{-12}.

Herlofson[19] made use of the above mentioned value of 10^{-2} for electrons created by moving atoms to deduce from theoretical considerations that the energy ratios of heat, light, and ionization produced by a meteor moving at 40 km/sec were 10^4; 10^2; 1 respectively. In the preceeding section, it was indicated that the ratio of the kinetic energy (nearly all of which appears as heat energy), to the ionization energy was of the order of 10^6 for a 60-km/sec meteor and 10^8 for a 20-km/sec meteor, which represent ionizing efficiencies several orders of magnitude smaller than that predicted by Herlofson. At the other extreme, Pierce[20] assumed that practically all of the kinetic energy was converted to ionization, though without offering any convincing evidence for the assumption.

It may be of interest to note that if the accepted values of atmospheric density adopted in this paper are reduced by a factor of five to ten, the computed values of k_3 will be increased by 10^2 to 10^3, and the rates of loss of mass will be similarly decreased. Better agreement would, then, be reached between the rates of electron production deduced from the loss of mass and the rates calculated from the radar observations. However, one would not presume to suggest that the present values of atmospheric density are too high on the slim experimental evidence of two radar meteors, especially when the values of some of the other factors are much more dubious.

IX. CONCLUSIONS

A statistical analysis of the meteor deceleration problem, using information from many observations, is required before any definite conclusions can be drawn, as the individual meteors may be expected to vary widely from the assumed mean characteristics. However, the data presented in this paper may be considered useful in showing that decelerations in the radar velocities do exist and can be measured (or else an alternative explanation of the apparent decelerations must be sought), and that the ionizing efficiency of the meteor can be determined if the atmosphere density is known. Conversely, if laboratory experiments or other observations could yield a reliable relationship for the ionizing efficiency, it should be possible to compute the density of the atmosphere at meteoric heights with a precision that may possibly be greater than the present photographic methods. If one accepts Lovell's formula, the radar parameters of the meteor can be determined fairly accurately, and furthermore, ρ depends on the one-sixth power of most of these parameters and is, therefore, remarkably insensitive to errors in them.

The desirability is pointed out of making further refinements in the radar and Doppler techniques of measuring velocities and decelerations, not only to increase the number of observations for statistical purposes, but to yield the rate of change of the deceleration from which the gradient of atmospheric density and the absolute temperature can be computed without reference to absolute signal strength or to an uncertain ionizing coefficient.

X. ACKNOWLEDGMENTS

The advice and cooperation of Peter M. Millman, Fred L. Whipple, and Barbara M. McKinley have been invaluable in the preparation of this paper. Special acknowledgment is due to E. L. R. Webb who, in addition to designing much of the equipment required in this work, was instrumental in developing the extension of the diffraction theory of radio reflections from meteor trails to include the case of a direct ground wave. Appreciation is expressed for the advice of B. G. Ballard, Director of the Radio and Electrical Engineering Division, NRC, who has also greatly facilitated the work by relieving the research staff of administrative duties.

[19] N. Herlofson, Rep. Prog. Phys. **11**, 444 (1948).
[20] J. A. Pierce, Phys. Rev. **11**, 88 (1947).

Latex Particle Size from X-Ray Diffraction Peaks* †

K. L. YUDOWITCH
Florida State University, Tallahassee, Florida
(Received August 31, 1950)

A high resolution camera was constructed for small angle x-ray scattering. Measurements with this camera on Dow latex 580G exhibit a number of concentric rings. Each ring is associated with a theoretical maximum of scattered intensity predicted by the spherical x-ray form factor corrected for packing.

The particle diameter is then computed from the position of each ring. Twenty measurements yield a diameter of 2780A with a probable error of less than ±1 percent.

THERE has been considerable interest in the determination of the particle size of Dow latex 580G, lot 3584.[1–5] Electron microscopists have been quick to adopt these remarkably uniform spheres as a magnification standard. In such applications, it is obviously most important that the correct diameter of these particles be accurately known. To this end, the values obtained independently in a number of laboratories have been tabulated.[4]

The measurements reported were obtained by six combinations of techniques, yielding diameters of 2520 to 3200A. It is significant that only four of the seventeen values tabulated are independent of electron microscopy. It was further indicated that the two light scattering values are not reliable. This leaves an ultracentrifuge value of 2520A±25A, a light microscope value of 3210A±240A, and an average of thirteen values utilizing electron microscopy of 2660A±80A. Neglecting the four largest values reported, a considered average of 2588A is reported for thirteen values in close agreement.

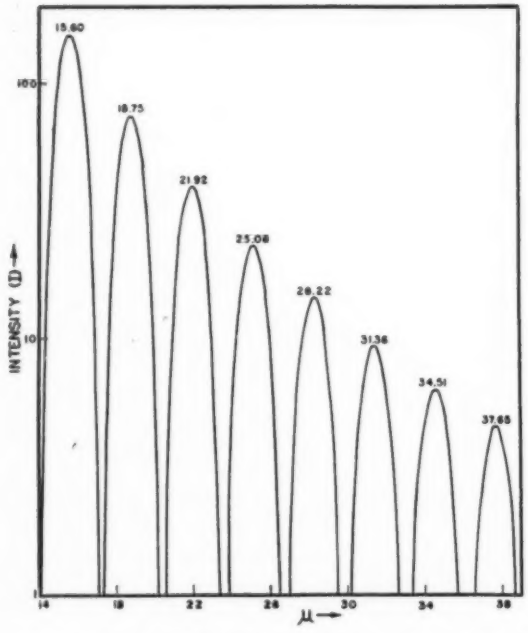


FIG. 1. Intensity distribution of x-rays scattered from packed uniform spheres.

The results of Kern and Kern[5] give strong indication of the necessity to correct these electron microscope values. They report that oblique shadow-casting increases the apparent diameter from 2590A±50A to 2840A±50A. It is suggested that this 6 percent deviation is caused by an artifact arising from the accumulation of electrons on the unshadowed particles.

The low ultracentrifuge value 2520A, the high light microscope value 3210A, the unshadowed electron microscope value 2590A, and the shadowed electron microscope value 2840A are in such poor agreement that the need for another independent determination became apparent at the 1949 Washington meeting of the Electron Microscope Society of America, where Gerould[4] and Kern and Kern[5] first reported their results.

A determination of the particle size by the method of small angle x-ray scattering was undertaken. The usual method of small angle x-ray scattering analysis utilizes the intensity variation over a very narrow range of angle close to the central beam. Due to the finite width of the central beam, and to the very narrow range over which the useful pattern is spread, this method has not been employed for particles larger than about 500A.

A further complication in the form of a deviation from the simple theoretical intensity distribution may be expected in the pattern of a specimen containing uniform particles which may be packed with some regularity. The intensity distribution from such a sample has not been theoretically predicted with very great accuracy. The approximate intensity distribution at small angles from such a sample has been discussed by Yudowitch.[6]

Of more interest in this case is the possibility of observing secondary intensity maxima. The positions

* This paper is based upon work sponsored, in part, by the Biological Department, Chemical Corps, Camp Detrick, Frederick, Maryland, under Contract No. W-18-064-CM-229 with the Florida State University.

† Presented at the Detroit meeting of the Electron Microscope Society of America, Sept. 14–16, 1950.

[1] R. C. Backus and R. C. Williams, J. Appl. Phys. **19**, 1186 (1948).

[2] R. C. Backus and R. C. Williams, J. Appl. Phys. **20**, 224 (1949).

[3] G. D. Scott, J. Appl. Phys. **20**, 417 (1949).

[4] C. H. Gerould, J. Appl. Phys. **21**, 183 (1950).

[5] S. F. Kern and R. A. Kern, J. Appl. Phys. **21**, 705 (1950).

[6] K. L. Yudowitch, J. Appl. Phys. **20**, 174 (1949).

of such peaks should give a more accurate determination of the particle size. For truly uniform spheres, it is expected that the series of secondary intensity peaks should be observable. Such peaks have been observed in the case of a fairly uniform gold sol.[6] The positions of these peaks are only slightly affected by the packing of the spheres. The distribution of scattered intensity is given, approximately, by Eq. 1.

$$I = MN^2\Phi^2(u)\left[1+P\left\{5\frac{\sin 2u}{2u}-6\Phi(2u)\right\}\right], \qquad (1)$$

in which I = intensity of x-rays, M = number of spheres in the sample, N = number of electrons per sphere, $\Phi(u)=3/u^3(\sin u - u\cos u)$, $u=\pi\epsilon D/\lambda$, ϵ = scattering angle, D = sphere diameter, λ = x-ray wavelength, and P = degree of packing.

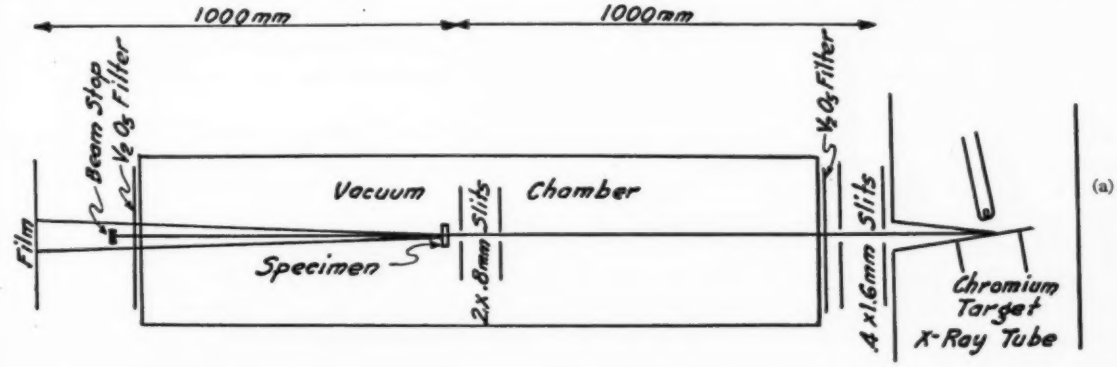
Assuming a reasonable value of the packing parameter ($P=0.5$), the intensity distribution predicted by this equation is shown in Fig. 1 for values of u from 14 to 39. This is well beyond the range of the small angle scattering which is usually analyzed ($u=0$ to 3).

In order to observe these closely spaced peaks, it is necessary to drastically reduce the collimation error. As the distance between peaks corresponds to a range of approximately π in the parameter u, it is obvious that the range of collimation error must be not much more than this value. In order to reduce the collimation error to about this value, a long evacuated camera was constructed as shown in Fig. 2. The collimating slits in this camera are of dimensions suitable to reduce the collimation error adequately for the observation of this pattern. In the construction of the camera the optimum collimation conditions[7] were utilized.

Chromium $K\alpha$-radiation (wavelength 2.286A) was selected as the longest wavelength readily available. As the absorption of this wavelength in a path of two meters reduces the intensity by a factor of about one thousand, the total path, except for a few centimeters on either end, was evacuated. The monochromatization was accomplished by filtering through vanadium pentoxide. A check with a crystal spectrometer indicated that the scattered radiation was essentially all chromium $K\alpha$. Any harder components present were apparently confined to the region of the main beam. Patterns were taken using a Machlett type A-2 chromium target diffraction tube operating self-rectified at 45 kilovolts and 15 milliamperes.

The specimen was prepared directly from the undiluted sample generously supplied by the Dow Chemical Company. A glass capillary of about ½-mm inside diameter was dipped into the latex. After allowing

[7] K. L. Yudowitch, J. Appl. Phys. **20**, 1232 (1949).



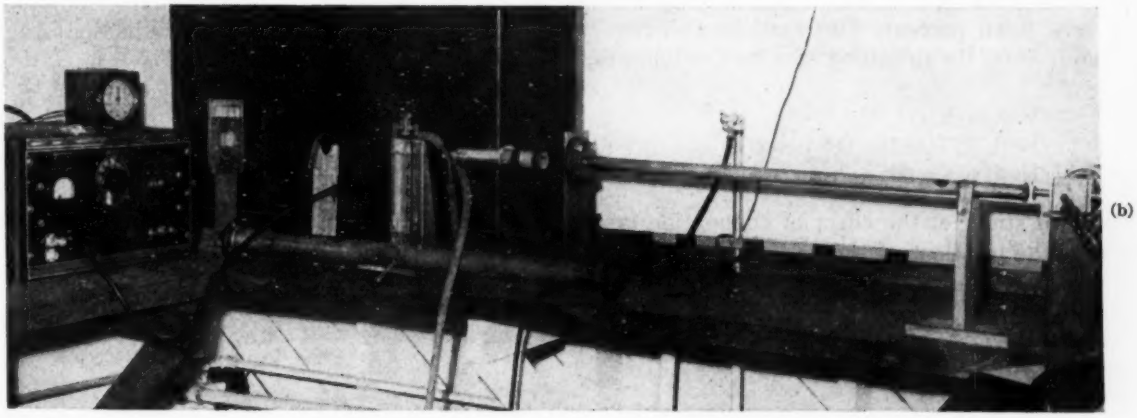


FIG. 2. (a) Sketch of high resolution small angle scattering camera, showing optimum collimation dimensions. (b) Photograph of high resolution small angle scattering camera, showing second portion of vacuum chamber removed to permit positioning of Geiger tube for slit alignment.

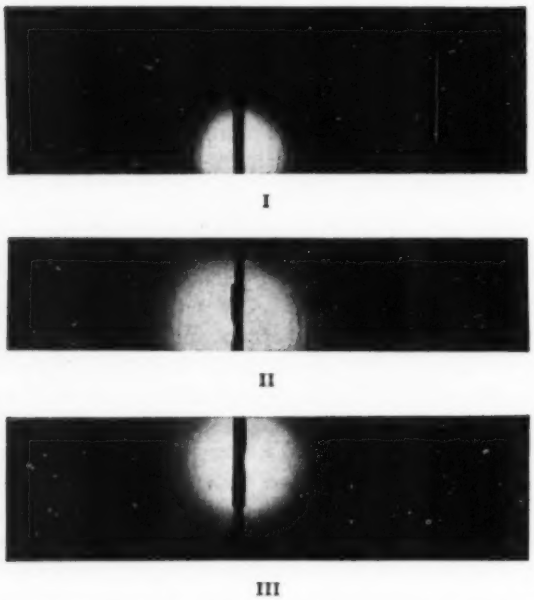


FIG. 3. Small angle x-ray scattering patterns from Dow latex 580G. I. 10-hour exposure. II. 30-hour exposure. III. 20-hour exposure.

several minutes for the latex to dry, it was protruded about one mm from the capillary. The protruded specimen was then put in place in front of the final collimating slit on the camera. Absorption measurement showed this thickness to be about a scattering optimum for the wavelength used. The resultant patterns of three exposures of 10, 30, and 20 hours are shown in Fig. 3. As many as a dozen concentric rings were observed in these negatives and in their densitometer records.

Several other patterns at shorter exposure times were taken. By comparing the short and long exposure patterns, it was possible, very simply, to identify each ring in the pattern as corresponding to a particular ring in the theoretical distribution of Fig. 1. The results of the densitometer records of the positions of these peaks appear in Table I. The mean value of (u/ϵ) is seen from these 20 measurements to have an average deviation of only ± 0.6 percent. The particle diameter is found simply from the definition of u by the following equation:

$$D=\lambda/\pi(u/\epsilon)$$
$$D=0.7277(3821\pm 0.6 \text{ percent}) \qquad (2)$$
$$D=2780\text{A}\pm 0.6 \text{ percent}.$$

A graphical study of the effect of collimation error reveals that the positions of the peaks may be shifted towards larger values of u. To obtain the collimation corrected curve, the intensity for each value of u in Fig. 1 is replaced by the mean value of the intensity for the interval $(u-0.5\delta u)$ to $(u+0.5\delta u)$. This "smearing" of the curve has little effect on the peak positions for $\delta u<\pi$, the approximate separation of adjacent peaks. Applying Eq. (8) of Yudowitch,[7] a value $\delta u=3.8$ is obtained for the geometry of this experiment. Due to imperfect slit alignment, the beam at the film measured only 0.54×3.00 mm instead of 0.80×3.20 mm as computed from the slits. Using these effective values for $2\theta_a$ and $2\theta_b$ in Eq. (8) yields an error $\delta u=2.9$. The "smearing" process indicated above has a negligible effect on the peak positions for $\delta u=2.9$. Hence, no correction to the 2780A figure is required for this cause.

This result is seen to be in agreement with the figure reported by Kern and Kern, and in disagreement with

TABLE I. Angular peak positions (ϵ) from patterns of Fig. 3, and corresponding values of (u/ϵ).

u	ϵ_I	u/ϵ_I	ϵ_{II}	u/ϵ_{II}	ϵ_{III}	u/ϵ_{III}
15.60	4.064	3839	4.090	3814	4.097	3808
18.75	4.901	3826	4.916	3814	4.863	3855
21.92	5.795	3783	5.706	3841	5.641	3886
25.08	6.543	3833	6.600	3800	6.534	3838
28.22	7.510	3758	7.406	3810	7.311	3860
31.36	8.347	3757	8.211	3819	8.225	3813
34.51			8.996	3836		
37.65			9.854	3821		

$(u/\epsilon)_{av}=3821\pm 0.6$ percent

the ultracentrifuge value and the considered average reported by Gerould. Barring variations attributable to specimen preparation and treatment, one might conclude that the effect reported by Kern and Kern is indeed substantiated. This conclusion would indicate that a value of 2780 to 2800A should be used rather than 2590A in utilizing the Dow latex as an electron microscopy magnification standard. It further supports Kern and Kern's suggestion that electron microscope images of unshadowed nonconducting objects may generally be subject to a spurious reduction in size.

CONCLUSION

The success of this experiment in resolving the diffraction rings and accurately determining particle size suggests the application of this method to the study of viruses and other uniform particles or molecules of diameter less than 0.3 micron.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

The Design of Molecular Pumps

ROBERT B. JACOBS*
Distillation Products Ind., Rochester, New York†
(Received September 5, 1950)

A theoretical treatment is given for the operating characteristics of molecular drag type pumps. The present work is based on kinetic theory and is limited to high vacuum operation. The solutions are simple and may be applied readily to practical pump designs. They lead to performance characteristics which are thought to be accurate to ±10 percent.

INTRODUCTION

THE molecular type of pump offers definite advantages over the high vacuum diffusion pump, for certain special applications. The complete absence of a pumping fluid in the molecular pump is perhaps its most outstanding point of superiority. This eliminates entirely the need for baffles or cold traps, as well as the possibility of "flashing" a pumping fluid into an evacuated system. The high compression ratios easily attainable with molecular pumps are also an advantage, making it possible to operate molecular pumps against forepressures of several millimeters of mercury.

Gaede[1] has given an excellent theoretical treatment of molecular pumps, based on hydrodynamical considerations.

The theory given here is based on kinetic principles and is considerably simpler. The present solutions are of such a form that they can be applied to actual pumps with far greater ease than to those of Gaede. Furthermore, they are adequate to predict the performance of complex pumps.

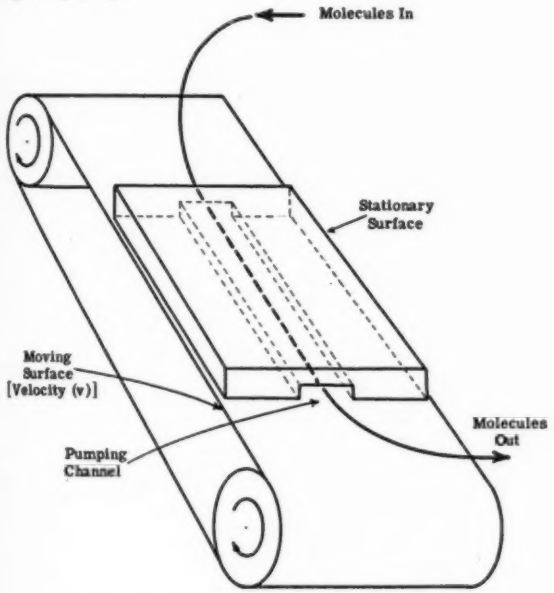


FIG. 1. Hypothetical molecular drag pump.

The pump designer is interested in predicting *compression ratios* and *pumping speeds* throughout the entire pumping range. He is further interested in knowing the requirements for machining tolerances. Finally, he would like to know for certain that *predicted* and *actual* performances do not differ by more than approximately 10 percent.

The following method, it is believed, achieves the above objectives. It is far simpler than that of Gaede, principally because it is based on "free molecular" considerations rather than on hydrodynamical principles. It is restricted, consequently, to dimension and pressures where collisions between the gas molecules are infrequent. A number of numerical calculations based on the present derivations have been made. These agree quite well with actual performance data on the pumps to which they have been applied.

DERIVATION OF PUMPING SPEED AND COMPRESSION RATIO (SIMPLIFIED FOR PUMPS OF CONSTANT PUMPING CROSS SECTION AND NO INTERNAL LEAKAGE PATH)

The action of a molecular pump can be understood by reference to Fig. 1. The schematic pump here considered consists of two parallel surfaces. One surface is in motion with a velocity (v), as indicated. The other surface is stationary and has a groove on the side facing the moving surface. The molecules between the plates are given a superimposed velocity, by the moving surface in the direction of its motion, each time they hit this surface. In consequence of this, there occurs a flow of molecules in this direction. This constitutes the pumping action.

Quantitatively, there will be an induced flow of velocity ($\bar{v}$) in the channel formed by the groove. This velocity is calculated as follows (assuming complete molecular accommodation):

$$\bar{v}=v\times\frac{\text{No. of molecules hitting moving surface}}{\text{No. of molecules hitting all surfaces}},$$

or, since all areas facing the channel are equivalent,

$$\bar{v}=v\times\frac{\text{area of moving surface (facing channel)}}{\text{total area of channel}},$$

* Now at Standard Oil Company (Indiana), Chicago, Illinois.
† Communication No. 175 from the Laboratories of Distillation Products Industries, Division of Eastman Kodak Company.
[1] W. Gaede, Ann. Physik **41**, 337 (1913).

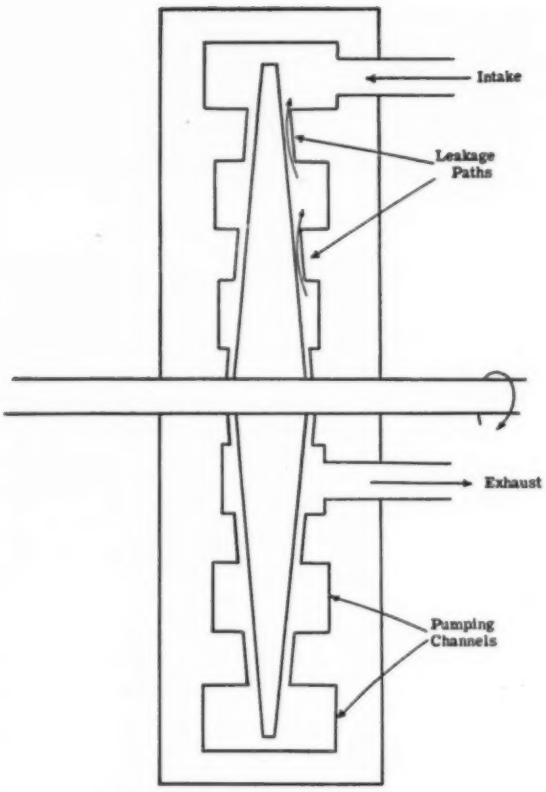


FIG. 2. Section of a disk type molecular pump (intake at periphery and exhaust toward axis).

or, more generally

$$\bar{v}=\sum v_i A_i/\sum A_i, \tag{1}$$

where different moving surfaces are involved simultaneously. In the above, all areas refer to those areas bounding the channel of flow.

Opposing this flow there will be a diffusion of molecules in the reverse direction in consequence of the pressure gradient established in this direction by the induced flow. This diffusion (volume/sec at unit pressure) is given by the following expression derived from kinetic theory.

$$Q_D=\frac{8}{3}\left(\frac{2}{\pi}\frac{p}{\rho}\right)^{\frac{1}{2}}\frac{A^2}{C}\frac{dp}{dl},\ddagger$$

whereas the induced flow may be written

$$Q_M=\bar{v}Ap. \tag{2}$$

It will be noted in passing that Q_M is the maximum pumping speed (i.e., the pumping speed at zero compression). The nomenclature used is as follows: A = area of pumping channel perpendicular to direction of flow, C = perimeter of pumping channel perpendicular to direction of flow, ρ = density of gas, p = pressure of gas, dl = element of channel length in direction of flow, v = velocity of moving surface, $\bar{v}$ = induced velocity of gas, Q_D = flow by diffusion (cc's/sec at unit pressure), Q_M = flow induced by moving surface (cc's/sec at unit pressure), Q = pumping speed = net flow = Q_M-Q_D, and $\alpha=Q/Q_M$.

If the diffusive flow is taken as some fraction (α) times the induced flow, the net flow (Q) will be as follows:

$$Q=(1-\alpha)Q_M. \tag{2a}$$

The compression is calculated by putting $Q_D=\alpha Q_M$, or

$$\frac{8}{3}\left(\frac{2}{\pi}\frac{p}{\rho}\right)^{\frac{1}{2}}\frac{A^2}{C}\frac{dp}{dl}=\alpha\bar{v}Ap,$$

or

$$\frac{dp}{p}=\frac{3\alpha\bar{v}}{16p}(2\pi p\rho)^{\frac{1}{2}}\frac{C}{A}dl,$$

whence, upon integration between the limits: l, 0 to L, and p, p_1 to p_2,

$$\ln\frac{p_2}{p_1}=\frac{3\alpha\bar{v}}{16p}(2\pi p\rho)^{\frac{1}{2}}\int_0^L\frac{C}{A}dl, \tag{3}$$

all terms outside the integral being constant.

When the cross section and radius of the pumping groove are constant along its length, and the pumping medium is air at room temperature, this becomes

$$\frac{p_2}{p_1}=\exp\left(\frac{\alpha\pi L\bar{v}}{2\cdot10^{-5}}\frac{C}{A}\right)\quad\text{(for cgs units).} \tag{3a}$$

Equation (3a) will give the relationship between the compression ratio (p_2/p_1) and the pumping speed if one makes use of Eq. (2a) to eliminate α.

Substituting Eq. (2a) in (3a), one obtains

$$\frac{p_2}{p_1}=\exp\left(\frac{Q_M-Q}{Q_M}\frac{\pi L\bar{v}}{2\cdot10^{-5}}\frac{C}{A}\right). \tag{3b}$$

This equation makes possible the calculation of the *compression ratio* (p_2/p_1) in terms of the physical *dimensions* of the pump and the *pumping speed* (Q). Accordingly, one may substitute into (3b) a series of arbitrary pumping speeds (subject to the condition that Q is always less than Q_M) and obtain a complete pumping curve (i.e., pumping speed *vs* compression ratio).

EXTENSION OF DERIVATIONS TO PRACTICAL PUMPS

The above development is applicable to an ideal pump, where internal leakages do not occur and where the pumping channel is of constant cross section. Actually, because of limitations of mechanical design, appreciable internal leakage occurs from the higher pressure regions of a pump toward the lower pressure regions. This leakage takes place across spaces of small dimensions which are provided as barriers between the adjacent pumping channels (see Fig. 2).

An internal leakage path may be considered as a small channel which allows back diffusion parallel to the main pumping channel. Then, since the back diffu-

‡ See, for example, L. B. Loeb, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York), p. 292.

TABLE I.

	Compression (meas.)	Compression (calc)	Pumping speed (meas.)	Pumping speed (calc)
(1)	10^{-3} mm-50×10^{-3} mm	same	73 liters/sec	68.5 liters/sec
(2)	—	10^{-3} mm-1 mm	not given	63.0 liters/sec
(3)	$6\times10^{-7}-2.3$ mm	same	not given	45 liters/sec

sion in the main pumping channel takes place according to the relation

$$Q_D=\frac{8}{3}\left(\frac{2}{\pi}\frac{p}{\rho}\right)^{\frac{1}{2}}\frac{A^2}{C}\frac{dp}{dl},$$

a similar expression may be used to calculate the leakage back diffusion by substituting values for A, C, and l which are appropriate to the leakage path. Or, since

$$Q_D \text{ (main channel only)}=k(dp/dl),$$

then

$$Q_L \text{ (leakage)}=k_1(dp'/dl')=k_2(dp/dl),$$

and if we put $k_2=\epsilon k$, we have

$$Q_D \text{ (total)}=(1+\epsilon)k(dp/dl).$$

Whence the corrected expression replacing Eq. (3a) becomes

$$\frac{p_2}{p_1}=\exp\left[\frac{\alpha\pi L}{2\cdot10^{-5}(1+\epsilon)}\frac{C}{A}\right] \tag{4}$$

In the above, dp'/dl' refers to the pressure gradient along the actual leakage path, which is usually related to the principal pressure gradient by some constant multiple, making the above transformation possible. Equation (4), then, may be used in a manner similar to Eq. (3) to obtain pumping characteristics.

In deriving Eq. (3), it was assumed that C/A and $\bar{v}$ were constant along the pumping channel. While this is true in the Gaede type pump, there are definite advantages in designing a pumping groove that tapers in the direction of the pumping action. Also, $\bar{v}$ may vary as a consequence of a shortening of the radius of the pumping groove as is the case in a disk type pump. For such pumps, we may take these factors into account by placing

$$\bar{v}C=v_0C_0(1-al) \quad \text{and} \quad A=A_0(1-bl),$$

whence the integral $\int_0^L(\bar{v}C/A)dl$, which was used in deriving (3a), now becomes $v_0C_0/A_0\int_0^L(1-al/1-bl)dl$, which upon integration gives $(v_0C_0/A_0)[\ln(1-bL)(a-b/b^2)+(a/b)L]$. Substituting in Eq. (3a), we obtain

$$\frac{p_2}{p_1}=\exp\left\{\frac{\alpha\pi}{(1+\epsilon)2\cdot10^{-5}}\frac{1}{b^2}\frac{\bar{v}_0C_0}{A_0}\times[(a-b)\ln(1-bL)+abL]\right\}. \tag{4a}$$

Equation (4a) is equivalent to (3a) but is the complete form for pumps with leakage, tapered pumping grooves, and a continuously changing pumping radius.

TABLE II.

Speed of rotation RPS	Compression Obs	Compression Calc	Compression Calc (Gaede)
78	4.9_7	4.9_5	5.0
104	6.9	6.6	7.0
182	12.4	11.6	12.8

USE OF EQUATIONS

Compression ratios (p_2/p_1) for practical designs are calculated by using Eq. (4a).

Pumping speeds (Q) will vary from zero to Q_M. Q_M is computed from Eqs. (1) and (2). For each *pumping speed* there will be a corresponding *compression ratio* (i.e., selection of a pumping speed determines α or Q/Q_M, which appears in Eq. (4a).

SOME NUMERICAL EXAMPLES

Friesen[2] has published dimensional and performance data on a pump of the general type illustrated in Fig. 2. The moving surface here is provided by a rotating disk. Two fixed surfaces are provided by a stator element into which tapered pumping grooves are machined. There are three parallel grooves of rectangular cross section in each stator half (i.e., similar to a triple screw). The direction of pumping is along a spiral, from the circumference to the axis. From the dimensions given by Friesen, the following numerical values for terms appearing in our Eq. (4a) are:

$$A_0=4.8_4 \text{ cm},$$
$$\bar{v}_0C_0=23{,}000 \text{ cm}^2/\text{sec},$$
$$(1+\epsilon)=1.05,\S$$
$$a=2.04\times10^{-3} \text{ cm}^{-1},$$
$$b=2.85\times10^{-3} \text{ cm}^{-1}.$$

The maximum measured pumping speed (73 liters/sec) is seen to be in good agreement with the speed (68.5 liters/sec) calculated for the compression at maximum speed. The pumping speed at maximum measured compression (6.7×10^{-7} mm to 2.3 mm) is not given, unfortunately. However, at this compression ratio the pumping speed is computed to have dropped off to 45 liters/sec, which appears reasonable. In all likelihood, since the pumping speed is dropping off rapidly at this compression ratio, equilibrium is reached between evolution of vapors and available pumping speed. In other words, the pump is just able to take care of the vapors evolved at 6×10^{-7} mm Hg.

Gaede[3] (see Fig. 3) gives the necessary dimensions on his pump for calculating compression and capacity. The Gaede pump is a multistage model, each stage being formed by a rectangular slot cut in the circumferential surface of the long cylindrical rotor. The stator accu-

[2] S. v. Friesen, Rev. Sci. Instr. **11**, 362 (1940).

§ Machining tolerances of the order of 0.005 cm were assumed in estimating $(1+\epsilon)$. In Table I calculated performance is compared with the available published data.

[3] W. Gaede, Ann. Physik **41**, 337 (1913).

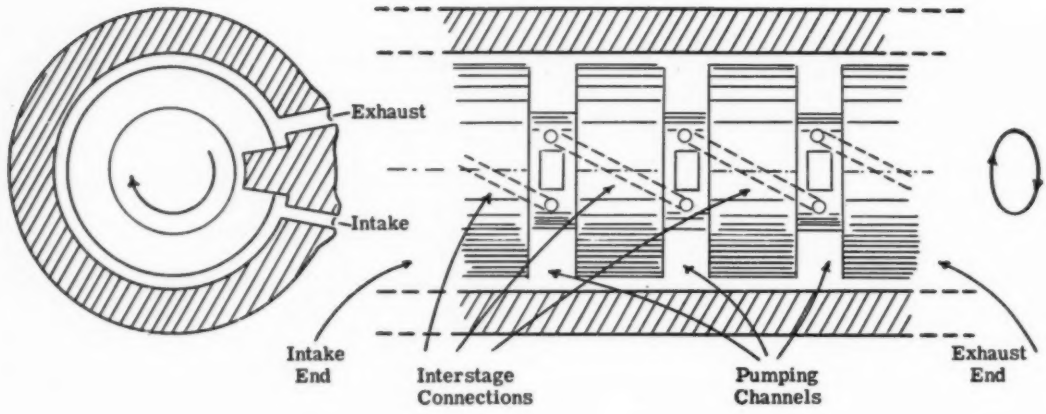


FIG. 3. Experimental pump of Gaede.

rately fits the rotor, and is fitted with wedge-shaped pieces which run in the rotor grooves, serving to separate the high pressure from the low pressure region of each stage of compression. Internal leakage thus takes place through two paths, i.e., from stage to stage parallel to the axis of the pump, and from the high pressure side to the low pressure side of a single stage around the wedge-shaped barrier which runs in the rotor groove. The Gaede pump has no taper. Accordingly, Eq. (3a) may be used to calculate compression. Observed values for compression are given only for the special case of no flow, (i.e., $\alpha=1$) a case for which the Gaede derivatives are considerably simplified. Accordingly, the following calculation will be restricted to the special case considered by Gaede. Principal data on the Gaede pump are as follows (for a single stage):

$\bar{v}=20.7n$ cm/sec (where n=revolutions/sec),
$\alpha=1.0$,
$L=23.5$ cm,
$C=6.2$ cm,
$A=1.5$ cm^2,
$1+\epsilon=1.54$.

Here, ϵ consists of two parts, ϵ_1 and ϵ_2, the former referring to stage leakage, and amounting to 0.40, and the latter being the leakage within a stage amounting to 0.14.

The above data yield the values in Table II for compression (per stage):

The calculated values are probably within the uncertainty of measurement. The greatest uncertainty in the data used are those upon which ϵ_1 and ϵ_2 are based, since these quantities depend principally upon the machining tolerances of the pump, which enter the calculations as squared terms. Fortunately, the values given by Gaede seem to be the result of careful measurements.

CONCLUSION

It is believed that the method herein presented provides a convenient and reasonably accurate means for predicting the characteristics of molecular pumps throughout their high vacuum range of operation. This is an advance over previous methods, which are too complex to be applied to practical pumps.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

Hypersonic Flow at a Mach Number of 10*

I. D. V. FARO, T. R. SMALL, AND F. K. HILL
Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland
(Received July 7, 1950)

Tests designed to produce supersonic streams at ten and sixteen times the local speed of sound are described. In contrast to the usual wind tunnel apparatus, a conical nozzle is used for the expansion section, and the supply pressure is varied throughout a wide range. Some properties of the flow at high Mach numbers are discussed, and, in particular, data are given showing a systematic dependence of Mach number on stagnation pressures, a phenomenon not previously reported.

THE production of supersonic streams in wind tunnels is usually accomplished by means of an expansion section in the form of a Laval nozzle; the Mach number of the stream depending primarily on the ratio of the exit area to the throat area. In order to obtain a uniform flow at the exit of the nozzle, the contour of the nozzle must be carefully designed and constructed. Though straightforward, practical methods

* Based on work done at the Applied Physics Laboratory, Johns Hopkins University, under Contract NOrd 7386 with the Bureau of Ordnance.

have been developed to determine the contour,[1–3] they still require skillful drawing and much calculation. In addition to this, the mechanical construction is usually difficult and the desired tolerances hard to maintain. The ratio of the exit area to the throat area changes very rapidly as the design Mach number, M, of the flow is increased. For large Mach numbers the throat may become so small physically that it becomes virtually a "source" producing irrotational, spatial flow. In such a case the design of the expanding nozzle can be greatly simplified. Based on this assumption, recent experiments have been made using a straight conical section to produce flow at ten times the local speed of sound.

The nozzle for which the results are herein reported consists of a 10-degree truncated cone with a small circular throat from which the gas expands and flows into a cylindrical section, the details and dimensions of which are shown in Fig. 1. The geometry has the following advantages:

(1) It permits a relatively small angular deflection of the stream at the throat and provides an essentially constant Mach number over every given cross section.
(2) It significantly reduces the length of the nozzle and thus minimizes the viscous effects between the stream and the wall surface.
(3) Since the nozzle shape is independent of the ratio of specific heats, γ, it may be used with any convenient working fluid, e.g., air, He, and CO_2.
(4) It may be quickly constructed to fine tolerances with ordinary machine shop tools without the necessity of any "hand" finishing.

The nozzle has the following disadvantages:

(1) The stream is circular in cross section, restricting the field of view for optical instruments; refer to Fig. 1.
(2) The test region in which the desired Mach number is essentially constant is limited by oblique shocks which arise at the juncture of the nozzle and test section.
(3) The small models necessitated by the over-all dimensions of the test section may give rise to significant scale effects.

None of these defects, however, preclude experimental studies on the properties of the flow.

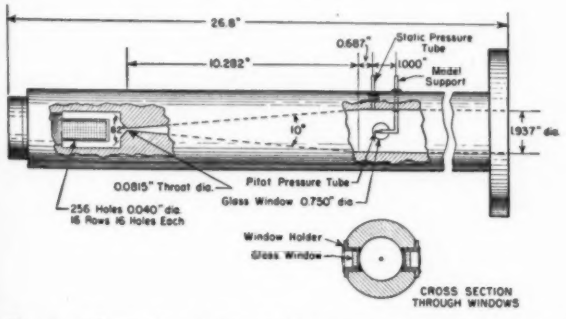


FIG. 1. Schematic of the working section of the hypersonic tunnel showing the settling chamber, conical nozzle, and cylindrical test section.

[1] H. W. Liepmann and A. E. Puckett, *Introduction to Aerodynamics of a Compressible Fluid* (John Wiley and Sons, Inc., New York, 1947).
[2] A. E. Puckett, J. Appl. Mech. **13**, A270 (1946).
[3] L. L. Cronvich, J. Aeronaut. Sci. **4**, No. 4 (1947).

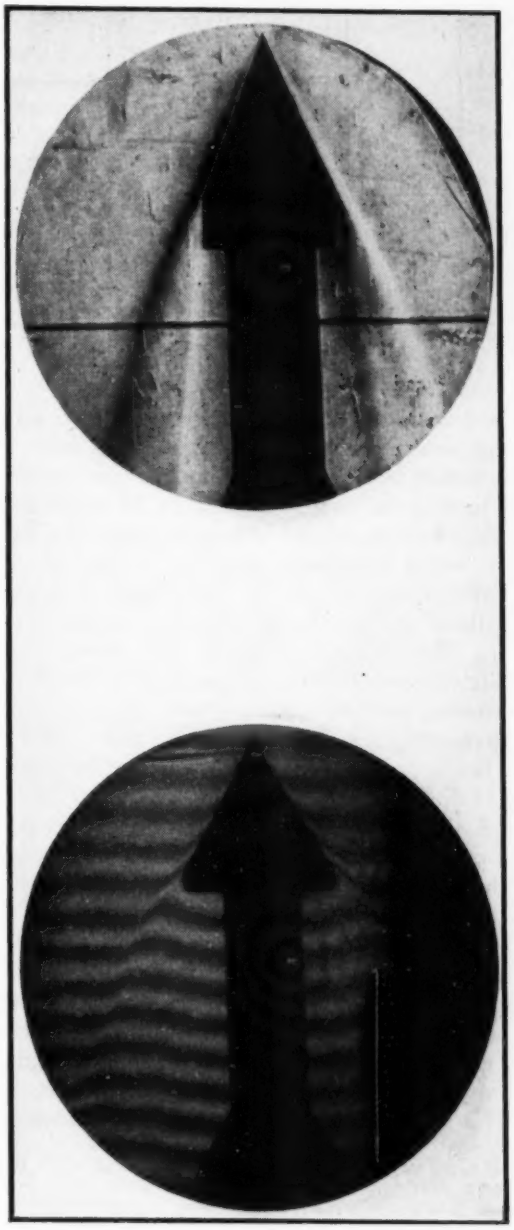


FIG. 2. Air flow at $M\sim10$: A Schlieren photograph of a 45° cone and an interferogram ($\lambda=5461$A) of a 60° cone.

Experiments with this nozzle and test section have been performed using air, nitrogen, and helium. Stagnation pressures, P_0, up to 3000 psi were used with air, and up to 2000 psi for the other gases. A suitable diffuser and vacuum reservoir provided the required compression ratio and both "blowdown" and continuous operations of the system were employed.

Results of the experimental work have shown that:

(1) This simplified nozzle gave flow close to the design Mach number; neither measurable nor visible flow irregularities have been noted to date. The Mach number determined by pitot head probes is 10.05 with a

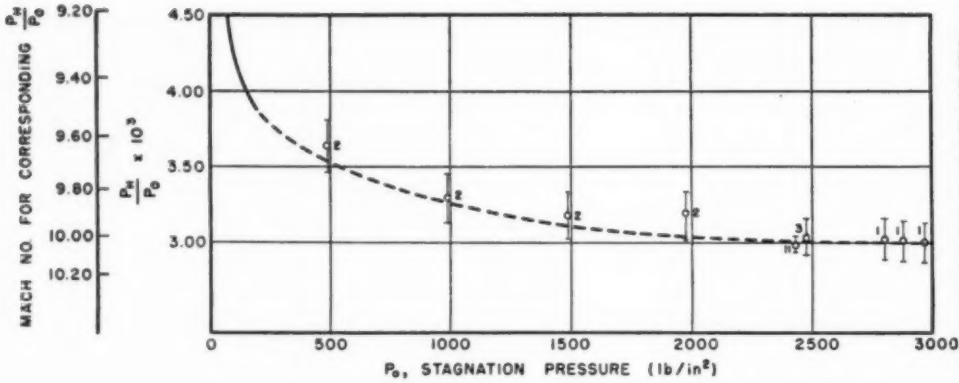


FIG. 3. Limits of data above $P_0=400$ psi given for total instrumentation and reading error, except at $P_0=2400$ psi. In this case ($P_0=2400$ psi) enough data were obtained to provide statistical weight and the 2σ-limits are shown (value of $P_H/P_0=2.99\times10^{-3}$ with $\sigma=0.014$). Figures next to the experimental points indicate the number of runs made. Solid line below $P_0=200$ psi represent data where the scatter is too small to show in this figure. Dashed line represents the estimated true behavior.

standard deviation of 2 percent for air.† Two other independent determinations of Mach number were made: (a) Schlieren photographs of shock waves from cones of 30° to 90° total angles (see Fig. 2). These gave only a statistical distribution of M values about $M=10$ from 7 to ∞, which is the expected result because of the low sensitivity of this method at high Mach numbers. (b) The ratio of test section density, ρ_1, to stagnation density, ρ_0. The stagnation density was computed from accurately known stagnation conditions of pressure and temperature, the test section density by interferometer measurements (Fig. 2). Tests were made with helium in the nozzle and the indicated Mach number was found to be 16.2 ± 0.2.

(2) A simple expanding diffuser provided 80 percent of the normal shock pressure recovery, thus indicating that the diffuser shape may not be as critical for hypersonic flow as it is for low supersonic channels.

(3) The Mach number was essentially constant for all values of $P_0>750$ psi with air, but showed a systematic decrease with lower values of P_0, see Fig. 3. This phenomenon, so far as we are aware, has not been noted elsewhere in the literature. A similar dependence of M on P_0 was observed with helium when $P_0<1000$ psi. This effect was determined by measurement of Pitot pressures and confirmed by interferometric measurements of the density in the test section. Although at low P_0 values the test section density and temperature are low and hence the kinematic viscosity relatively high, the available theories for boundary layer formation do not account for this change in Mach number. A slight increase (3 percent) in Pitot head pressure was observed when the water content of the air was increased by a factor of 2.6. However, we were not able to account for the magnitude of the systematic variation of M at low values of P_0, in terms of viscosity, humidity, or other known thermodynamic properties of the flow. Although heat transfer could produce this change in M at low densities, neither radiation nor conduction provide an adequate mechanism; convective processes are under study at the present time.

(4) Static pressure measurements by means of a probe and wall taps indicated a value of M about 1 Mach number less than the Pitot pressures. The low static densities (static test section pressures $\sim$3 mm Hg or less) are believed to give rise to slip flow when the gas passes over a fixed surface, so that the measured values of static pressure on a wall or probe may not be true indications of the free stream static pressure. This effect has been observed and reported by other investigators,[4] and presents an important problem which needs to be solved for a complete understanding of the properties of hypersonic flow.

It is planned to continue these studies of the properties of streams at high Mach numbers and to report the results in more detail at a future date.

† *Note added in proof:*—For supply temperatures below 1000°R, free stream static pressures are somewhat higher than expected (as stated in 4 under results) on the basis of other M determinations. Recent experiments with nitrogen at stagnation temperatures up to 2000°R have resulted in lowering the static pressure to values consistent with the other measurements. This indicates some condensation at supply temperatures below 1000°R, but the onset of condensation apparently occurs at temperature-pressure conditions more favorable than those expected from vapor-pressure data.

[4] Symposium on hypersonic flow and low density flow, Phys. Rev. **79**, 546 (1950).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 2 FEBRUARY, 1951

Calculation of the Electrical Capacitance of a Cube

DANIEL KINSETH REITAN AND THOMAS JAMES HIGGINS
University of Wisconsin, Madison, Wisconsin
(Received August 18, 1950)

The basic theory of calculation of the capacitance of a given geometrical configuration by the use of subareas is advanced and applied to solve the long-standing problem of the accurate evaluation of the capacitance C of a cube of side a. The best previously published determination is $0.62211a<C<0.71055a$. The value obtained of $C\approx 0.655a$ esu is both a lower limit and very close to the exact value.

I. INTRODUCTION

IN a series of recent, interesting papers, Polya and Szegö[1–4] have discussed the problem of the approximate determination of the electrical capacitance of a specified solid, have advanced the theory of several different approximate methods of calculation, and have applied this theory to determine approximate values of capacitance for a number of specific solids. In particular, they[3] advance upper and lower bounds for the capacitance of a cube of side a: Thus, $0.62211a<C<0.71055a$ esu. It appears that these limits are the best published values for the capacitance of a cube, exact determination of which has long comprised one of the major unsolved problems of electrostatic theory.*

It is the purpose of this paper to advance the theory of a certain approximate method of calculation which enables determination, to any desired degree of accuracy, of the capacitance of a cube, and to calculate by this method, a value for the capacitance which is very close to the exact value.

II. BASIC THEORY

The "method of subareas" is generally applicable to approximation of the capacitance and charge distribution of a body of arbitrary geometry. It was employed by Maxwell[5] to determine the capacitance of a square area and recently by the authors[6] to determine the capacitance and charge distribution of a plane circular annulus. The essential theory is to be epitomized as follows: Let the surface area of the given body be subdivided into n subareas, $A_i (i=1, \cdots, n)$ which are:

(1) Of such small area by comparison with the given area that the charge density q_i is essentially constant over each subarea A_i.

(2) Of such shape that assumption of uniform charge density enables simple calculation of the potential V_i produced by this uniformly distributed charge.

(3) Of such dimensions and shape that if the subarea A_i were alone in space, the potential V_{ij} produced by A_i over that part of space, which is actually occupied by another subarea A_j, is essentially constant; and similarly for the potential V_{ii} produced by q_i over A_i itself.

Under these assumptions calculation proceeds as follows: By (1), V_{ij} produced over A_j by the uniform charge distribution q_i on A_i is proportional to q_i, whence $V_{ij}=k_{ij}q_i$. Hence by (1) and (3) the total potential over

$$A_j \text{ is } V_j=\sum_{i=1}^{n} V_{ij}=\sum_{i=1}^{n} k_{ij}q_i,$$

a linear equation in the n unknowns $q_i (i=1, \cdots, n)$. Proceeding thus to form the total potential over each subarea yields the set of n equations

$$V_j=\sum_{i=1}^{n} k_{ij}q_i \quad (j=1, \cdots, n).$$

A well-known theorem in electrostatic theory states that the potential is constant, $V=V_0$, over a charged

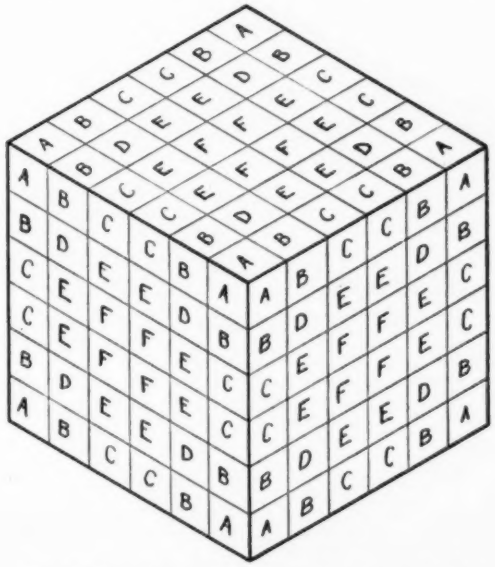


FIG. 1. Distribution of subareas on faces of cube.

[1] G. Polya and G. Szego, Am. J. Math. **67**, 1–32 (1945).
[2] G. Szego, Bull. Am. Math. Soc. **51**, 325–350 (1945).
[3] G. Polya, Am. Math. Monthly **54**, 201–206 (1947).
[4] G. Polya, Quart. Appl. Math. **6**, 267–277 (1948).
* It was related by Kirchhoff, that shortly before his death Dirichlet solved the problem of the distribution of electricity on a rectangular parallelepiped. If so, the solution has been lost.
[5] J. C. Maxwell, *A Treatise on Electricity and Magnetism* (Oxford University Press, Oxford, England, 1892), third edition, Vol. 1, 148–154. H. Cavendish, *Electrical Researches* (Cambridge University Press, Cambridge, England, 1879, edited by J. C. Maxwell), 426–427.
[6] T. J. Higgins and D. K. Reitan, Trans. Am. Inst. Elec. Engrs. (publication pending).

TABLE I.

Number of subareas	Capacitance of unit cube
6	0.437 esu
24	0.532 esu
96	0.648 esu
216	0.6555 esu

conductor whereon the charge is in equilibrium. Imposing this condition over each of the subareas yields

$$V_0=\sum_{i=1}^{n} k_{ij}q_i \quad (j=1, \cdots, n),$$

a set of n linear, algebraic equations in the n unknowns $q_i(i=1, \cdots, n)$. Solution of this set of equations yields a set of values for the uniform charge densities q_i over the subareas. Accordingly, the approximate value of the total charge Q is

$$Q=\sum_{i=1}^{n} q_iA_i.$$

Finally, the approximate value of the capacitance follows from $C=Q/V_0$ (the q_i, and hence Q, are expressed in terms of V_0, which cancels out in taking the ratio). Knowledge of the approximate distribution of charge density over the area A is yielded by the known values of the q_i.

It appears then that the capacitance and charge density can be evaluated to the degree of accuracy desired by taking subareas of sufficiently small size. Of course, the labor involved in solving the set of n linear, algebraic equations increases rapidly with n; but in many geometrical shapes of particular interest the approximation obtained with small n is surprisingly good; as evidenced, for example, by the curve of Fig. 2.

III. THE CAPACITANCE OF A CUBE

Consider a cube whose surface area is divided into 216 unit subareas, 36 unit subareas per face (Fig. 1). Symmetry limits the number of unknowns to 6. Now the potential P_A existing at the center of subarea A is comprised of that contributed by the charge upon subarea A plus that contributed by all of the other subareas on the faces of the cube. Consider the charge distribution on each subarea as uniform. Then as detailed in Appendix I, the potential at the center of a unit square subarea due to its own charge is $3.52549q$, where q is the charge of the unit square; the potential at the center of the unit square due to the charge on any other unit square is well approximated by q'/d, where q' is the charge on the other unit square and d is the distance between the centers of the two squares. Utilizing these relationships to determine the total potential at the center of subarea A, and performing the indicated arithmetic, yields the first of the following set of equations. Proceeding similarly for subareas B to F yields the remaining equations.†

$$V_A=9.5564q_A+11.7114q_B+10.0497q_C+5.3803q_D+10.0610q_E+4.9481q_F,$$
$$V_B=5.8559q_A+14.4892q_B+11.3004q_C+5.9485q_D+11.1248q_E+5.4322q_F,$$
$$V_C=5.0246q_A+11.3011q_B+15.2190q_C+5.7892q_D+12.0741q_E+5.7785q_F,$$
$$V_D=5.3803q_A+11.8970q_B+11.5788q_C+8.6750q_D+12.6493q_E+6.1879q_F,$$
$$V_E=5.0301q_A+11.1246q_B+12.0741q_C+6.3252q_D+15.7544q_E+6.9904q_F,$$
$$V_F=4.9479q_A+10.8652q_B+11.5572q_C+6.1879q_D+13.9789q_E+10.6362q_F.$$

Imposing the condition that $V_A=V_B=V_C=V_D=V_E=V_F=V_0$ yields:

$$q_A=0.03015V_0, \quad q_D=0.01325V_0,$$
$$q_B=0.02145V_0, \quad q_E=0.01273V_0,$$
$$q_C=0.02026V_0, \quad q_F=0.01159V_0.$$

The total charge upon the cube is:

$$Q=24q_A+48q_B+48q_C+24q_D+48q_E+24q_F=3.9328V_0.$$

Hence: $C=Q/6V_0=0.6555$ esu $=0.7283$ $\mu\mu f$. Accordingly, in virtue of the well-known theorem of electrostatics, that the capacitances of geometrically similar solids are in proportion to the ratio of corresponding linear dimensions, it follows that the approximation to the capacitance of a cube of side a is:

$$C=0.6555a \text{ esu}=0.7283a\ \mu\mu f.$$

Confirmatively, this value of $0.6555a$ esu falls within limits obtained by Polya and Szego, as cited in Sec. I. Moreover, as evidenced by the values of Table I and the corresponding curve of Fig. 2, a value of $0.655+a$

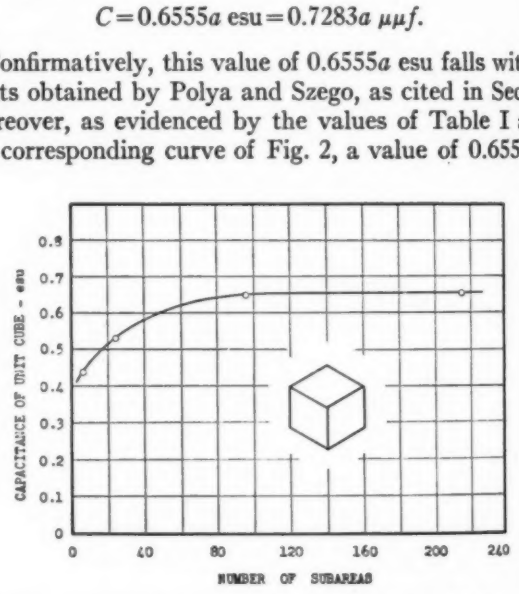


FIG. 2. Approximate capacitance *versus* number of subareas.

† In virtue of the geometric symmetry of the cube the coefficients k_{ij} and k_{ji} $(i\neq j)$ should either be equal, or one ought be twice the other. The corresponding coefficients of this set of equations satisfy this condition, except for slight differences stemming from independent calculation of each of the paired coefficients. These slight differences affect, at most, the last digit of $C=0.6555$ in Table I by about ± 1 in the last digit.

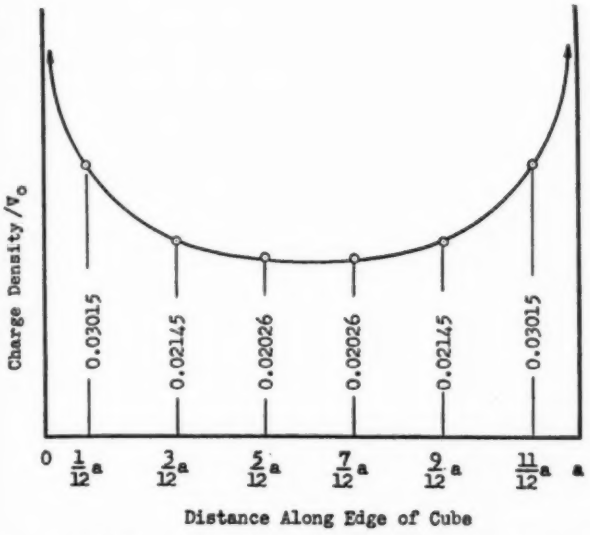


FIG. 3. Variation of charge density along a line drawn through centers of subareas comprising any outermost row.

esu is both a lower limit and very close to the unknown exact value of the capacitance of a cube of edge a.

IV. CHARGE DISTRIBUTION

The tabulated values of $q_A, \cdots, q_F$ are good approximations to the actual charge density at the centers of the corresponding subareas. By well-known theory, the charge density at the sharp edges of the cube is infinite. If, then, one face of the cube is chosen as reference; ordinates of finite height erected at the center of every subarea and of infinite height at the center of each side of a subarea which is coincident with an edge of the cube; the heights of the finite ordinates made equal to the values of the charge density at the corresponding points; and the tops of all ordinates joined by a smooth surface, the surface thus obtained affords a good approximation to the charge density at any point in the surface of the cube. Figure 3 is a plot of the intersection of this surface and of a plane perpendicular to the reference face of the cube along the center line of an outer row of subareas A, B, C, C, B, A, of Fig. 1.

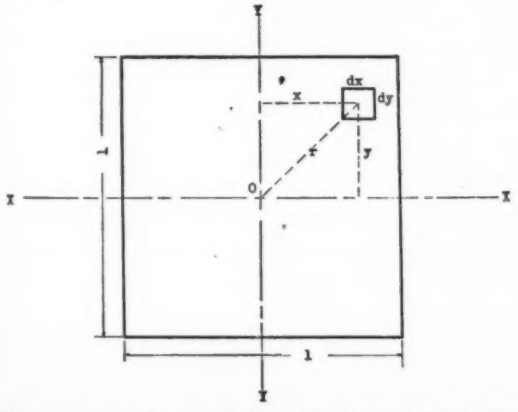


FIG. 4. Coordinate system for calculation of potential at center of unit square.

V. SUMMARY

1. The basic theory of approximate calculation of the capacitance of a given geometrical configuration by the use of subareas is advanced and applied to the long-standing problem of the accurate evaluation of the capacitance of a cube.

2. Approximate calculation of the capacitance, using 216 subareas, yields a numerical value of $C \approx 0.655a$ esu which is both a lower limit and (as evidenced by Fig. 2) very close to the exact value.

3. Additionally, the calculation involved in determining C enables accurate knowledge of the distribution of charge over the surface of the cube, as indicated in Fig. 3.

4. In general it may be noted that in various problems of practical interest, knowledge is required of the capacitances of geometrical configurations which are of such shape as to be intractable to exact calculation because of the mathematical difficulties involved, yet which can be easily evaluated by the method of subareas. Thus the computation of the performance of the paddle viscometer with rectangular blades (material testing), of the flow of fluid through a rectangular or annular orifice in a plane barrier (hydrodynamics), of the motion of a rectangular or annular plate through an infinite stationary fluid (aerodynamics), or the effects of annular guard rings in a precision standard capacitor (electrical measurements) requires knowledge of a parameter equal in magnitude to that of the capacitance of the corresponding rectangle or circular annulus. Exact formulas for the capacitances of these configurations have not been determined. However, the method of subareas enables very accurate calculation of the desired capacitances.[6,7]

APPENDIX I

Assume that the unit square of Fig. 4 is uniformly charged to a total value q. The charge density at any point on the square is

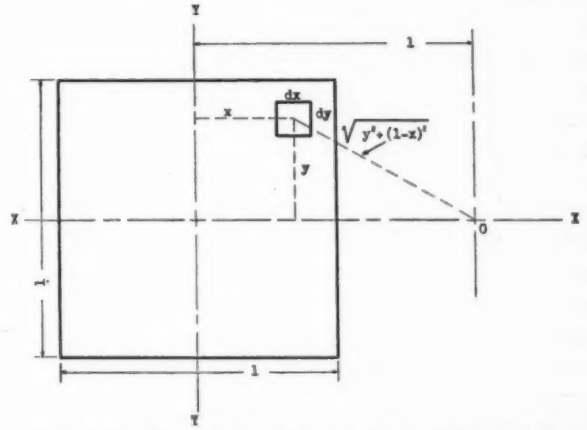


FIG. 5. Coordinate system for calculation of potential at center of adjacent unit square.

[7] D. K. Reitan, "An approximate calculation of the electrical capacities of rectangular and annular areas," M.S.E.E. thesis, University of Wisconsin (1949).

$q/\text{area}=q/1=q$. The total charge on an incremental area $dxdy$ is $qdxdy$. Each increment of charge $qdxdy$ contributes to the potential at point 0, a distance r from it, an increment of potential $dV=qdxdy/r$. Accordingly, the total potential V at 0 is after integration by Dwight 200.01, substitution of u^{-1} for $2y$, and a second integration by Dwight 731.2,

$$V=4q\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}}\frac{dxdy}{(x^2+y^2)^{\frac{1}{2}}}=4q\int_0^{\frac{1}{2}}\sinh^{-1}\frac{1}{2y}dy=2q\int_1^{\infty}u^{-2}\sinh^{-1}u\,du$$
$$=4q\ln[1+(2)^{\frac{1}{2}}]=3.52549q.$$

APPENDIX II

Assume that the unit square of Fig. 5 is uniformly charged to a total value q. The total charge upon the incremental area $dxdy$ is $qdxdy$. Each increment of charge $qdxdy$ contributes to the potential at point 0, an increment of potential $dV=qdxdy/(y^2+(1-x)^2)^{\frac{1}{2}}$. Accordingly, the total potential V at point 0 is:

$$V=2q\int_{x=-\frac{1}{2}}^{\frac{1}{2}}\int_{y=0}^{\frac{1}{2}}\frac{dydx}{(y^2+(1-x)^2)^{\frac{1}{2}}}=-2q\int_{z=\frac{1}{2}}^{\frac{1}{2}}\int_{y=0}^{\frac{1}{2}}\frac{dydz}{(y^2+z^2)^{\frac{1}{2}}}$$
$$=-2q\left[z\ln\left(\frac{1}{2z}+\left(\left(\frac{1}{2z}\right)^2+1\right)^{\frac{1}{2}}\right)+\tfrac{1}{2}\ln(2z+((2z)^2+1)^{\frac{1}{2}})\right]^{\frac{1}{2}}$$
$$=1.031q=q/0.970.$$

This calculation indicates that the numerical value of the potential produced by a uniformly charged square, at a point located a distance equal to or greater than unit distance from the center of the square, is approximately equal to the numerical value of the charge on the square divided by the distance to the point considered.

The Properties of Evaporated Gold and Tungsten Oxides

P. G. WILKINSON
U. S. Naval Research Laboratory, Washington, D. C.
(Received August 2, 1950)

Gold blacks, prepared by evaporating gold from a tungsten filament in tank nitrogen at several mm pressure, were found to contain tungsten oxides which greatly affected the optical and electrical properties. Deposits containing more than 0.5 mole percent of tungsten oxides were black and showed resistivities 10^3 to 10^8 times that of bulk gold. The resistivities of gold blacks increased on exposure to oxygen. On the other hand, oxide-free deposits were yellow and possessed a low stable resistivity, unaffected by exposure to oxygen. Exposure to air or oxygen caused an increase in the electrical resistance, and optical transmission of films of tungsten dioxide and tungsten trioxide evaporated in high vacuum. It was concluded that evaporated films of pure gold are unaffected by exposure to air and that the changes in optical and electrical characteristics, often observed in evaporated gold films exposed to air, may be attributed to contamination by oxides of tungsten. It was also concluded that so-called gold blacks are black because of the presence of tungsten oxides, since gold evaporated in pressures of several millimeters of oxygen-free nitrogen formed bright reflecting films rather than black absorbing deposits.

INTRODUCTION

HIGH vacuum evaporated gold films are metallic in appearance and, immediately after preparation, show a fall in electrical resistance due to a decay of lattice defects introduced in the condensation process.[1,2] On the other hand, gold deposits formed by evaporation in a poor vacuum or in a gas such as unpurified tank nitrogen are no longer metallic but black in appearance and are good absorbers of radiation. These are, therefore, known as "blacks." Gold blacks are now almost universally used as absorbing coatings on thermocouples and bolometers.* The term "smokes," as used in this paper, will include both blacks and brownish-yellow deposits, both types being formed in a gas atmosphere. The blacks and some of the brown deposits prepared in this way show no resistance decrease when exposed to air, but, instead, showed an increasing resistance of the order of megohms. This effect has also been noted for evaporated deposits of silver,[3] lead,[4] and palladium, platinum, rhodium, and iridium.[5]

Harris, McGinnes, and Siegel[6] have found considerable differences in the optical transmission of gold blacks prepared by heating a tungsten filament coated with gold in nitrogen containing oxygen in amounts varying from 0.3 to 5 percent. The deposits formed in the 0.3 percent oxygen and nitrogen mixtures showed low optical transmission from 0.4μ to 15μ, while those prepared in higher percentages of oxygen showed increased transmission beyond 2μ.

According to Langmuir,[7] tungsten heated above 700°C reacts heterogeneously with low pressure oxygen to form the volatile tungsten trioxide, while tungsten heated in oxygen-nitrogen mixtures forms an oxide of the stoichiometric formula W_2O_5. Nitrogen, however, reacts only with tungsten vapor, which is present at a negligible pressure below 2300°C.[7] Tungsten oxides, formed by heating a tungsten filament to 1500°C in 2 microns of flowing oxygen, have been used to shadow

[1] V. Vand, Proc. Phys. Soc. (London) **55**, 222 (1943).
[2] P. G. Wilkinson and L. S. Birks, J. Appl. Phys. **20**, 1168 (1949).
* Formerly, absorbing coatings frequently used were bismuth and zinc blacks whose properties have been described by A. H. Pfund, Rev. Sci. Instr. **1**, 397 (1930); J. Opt. Soc. Am. **23**, 375 (1933).

[3] P. Morozov and M. Butslov, J. Tech. Phys. (USSR) **16**, 857 (1946).
[4] E. Armi, Phys. Rev. **63**, 451 (1943).
[5] M. AuWärter and K. Ruthardt, Z. Elekt. **44**, 579 (1938).
[6] Harris, McGinnies, and Siegel, J. Opt. Soc. Am. **38**, 582 (1938).
[7] I. Langmuir, J. Am. Chem. Soc. **35**, 105, 931 (1913).

electron microscope specimens.[8] Various oxides of tungsten, with a stoichiometric formula corresponding approximately to W_4O_{11}, have been investigated by x-ray diffraction,[9] but the existence of a true molecule with this formula is not considered established. The crystallographic structures have been said to bear a resemblance to those of both WO_2 and WO_3.[10]

It is thought that all of the gold blacks measured by Harris probably contained oxides of tungsten formed by simultaneous oxidation of the tungsten filament during the gold evaporation. Evidence is presented in this paper to show that the optical and electrical properties of gold smokes are modified largely by impurities such as tungsten oxides, and that if they are not present in the deposit, it is a yellow high reflecting film, whether evaporated at high or low pressure.

SPECIMEN PREPARATION AND TESTING

The evaporations of gold were carried out in a bell jar modified to allow purified tank nitrogen† to flow through it at a constant rate. The gold was evaporated from a conical tungsten wire cup electrically heated and maintained at constant temperature by observation with a Leeds and Northrup Optical Pyrometer. In some evaporations, tungsten dioxide and tungsten trioxide were vaporized from a quartz cup fitted into a tungsten heater coil.

The films and smokes were deposited on electron microscope specimen screens coated with Formvar and on microscope slides silver tipped to permit measurements of electrical resistance. These were located at a distance of 7 cm from the tungsten cup. The clear area on the microscope slides measured 1.8 cm by 2.5 cm. The surface density of the deposit was determined by weighing a measured area with a microbalance.

Some later evaporations were made in an all Pyrex evaporator constructed from a five-liter flask, which could be readily evacuated to 10^{-6} mm pressure and to which was attached a gas purification train, storage vessels, manometers, liquid-air traps, and a gas analysis system. The electrical resistance slides were located below and in thermal contact with a glass finger sealed into the flask, which permitted cooling of the specimen with refrigerant if desired. The clear area of these electrical resistance slides measured 1 cm by $1\frac{1}{2}$ cm and were located 7 cm above the filament. This apparatus allowed evaporations to be carried out in highly purified flowing nitrogen with no danger of contamination. The pressures of nitrogen employed were 10^{-6}, 10^{-3}, 1, 5, 10, and 13 mm of mercury.

The oxygen content of the tank nitrogen was estimated by bubbling the gas through freshly precipitated manganous hydroxide. The $MnO(OH)_2$ released was then caused to displace an equivalent quantity of iodine which was then titrated with 0.01*N* sodium thiosulfate.

Tungsten trioxide was prepared from sodium tungstate by the method of Schumb and Hartford.[12]

Tungsten dioxide was prepared by heating the trioxide in vacuum for 15 hours with tungsten powder at 1000°C.

The gold used was in the form of 30-mil wire and was obtained from J. Bishop and Co. Platinum works. It analyzed about 0.03 percent silver as the major impurity.

GOLD DEPOSITS CONTAINING TUNGSTEN OXIDES

By controlling the evaporation conditions, gold smokes may be prepared containing various percentages of tungsten oxides. The Pyrex evaporator constructed in this laboratory permitted the preparation of gold smokes in nitrogen containing as low as 0.001 percent oxygen. Gold smokes prepared in the presence of more than about 0.1μ-pressure of oxygen always showed contamination with tungsten on spectrochemical analysis. With the greatest care being taken to insure purity of gas and gold, gold smokes were prepared in 1 mm of flowing nitrogen, which were no longer black but possessed high reflectivity and a stable resistivity approaching that of bulk gold and were unaffected by exposure to oxygen. No tungsten was detected in these deposits. Other samples of gold smokes containing a few percent copper and silver and no tungsten were also black in visual appearance. Blacks were also made by evaporating mixtures of tungsten trioxide and gold, as well as molybdenum trioxide and gold in highly purified nitrogen.

The Chemical Composition of Gold Blacks

The percentage of tungsten found in a gold smoke, prepared by evaporation in impure nitrogen, depends entirely on the evaporation conditions. As has been shown previously,[7,8] tungsten heated in even small partial pressures of oxygen forms oxides which readily evaporate. The rate of formation of tungsten oxides, at a constant filament temperature, depends on the rate at which molecules of oxygen strike the filament, which, in turn, is determined by the partial pressure of oxygen and the rate at which oxygen is passed through the evaporation chamber. Increasing pressures of nitrogen would be expected to decrease the rate of tungsten oxide formation by competing with oxygen for active centers on the filament. Also, the relative number of tungsten oxide molecules and gold atoms which reach the collecting surface would be expected to depend on

[8] P. G. Wilkinson and L. S. Birks, NRL Report P-3443 (1949); J. Appl. Phys. **21**, 60 (1950).
[9] E. Tarjan, Naturwiss. **19**, 166 (1931). F. Ebert and H. Flaschi, Z. anorg. u. allgem. Chem. **217**, 95 (1934). G. Hagg and A. Magneli, Arkiv. Kemi, Mineral. Geol. **19A**, 2 (1944).
[10] A. Magneli, Arkiv. Kemi Mineral. Geol. **1**, 223 (1949).
† The tank nitrogen was purified by it passing over copper gauze heated to 500°C and phosphorus pentoxide at a rate of 2 liters per hour. The resulting gas analyzed 0.001 percent oxygen and was stored above atmospheric pressure in previously evacuated 5-liter flasks.

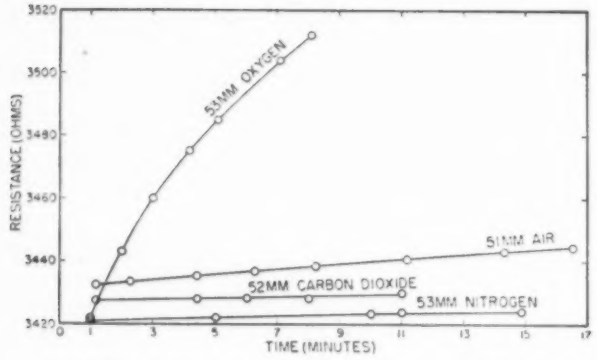


FIG. 1. The effect of atmospheric gases on the electrical resistance of a gold deposit condensed in one micron of air. Rate of deposition is one $\mu g/cm^2/min$. Thickness is 34 $\mu g/cm^2$.

the total pressure (mostly nitrogen), since these substances are carried along by convection currents of gas. Increasing the tungsten filament temperature raises the evaporation rate of gold in proportion to $T^{-\frac{1}{2}}\exp(-B/T)$ and the rate of oxide formation in proportion to $\exp(-A/T)$, where A and B are constants and T is the absolute temperature. Finally, the ratio of quantities of tungsten oxide to gold which are evaporated depends on the free surface area of the evaporating gold and the exposed surface area of the tungsten filament.

By maintaining the total pressure at a value between 1 and 13 mm and the ratio of oxygen to nitrogen at 5×10^{-3}, a number of large samples of blacks were prepared and analyzed by spectrochemical methods. The following equation was found to hold (within 20 percent):

$$\text{Percentage tungsten oxides} = 10/G, \qquad (1)$$

where G is the rate of deposit in micrograms per square centimeter per minute and was varied by changing the filament temperature (from 1190°C to 1430°C). Using this equation, the composition of small black samples prepared under like conditions could be estimated.

Resistance-Time Studies

The preliminary resistance-time studies were made on bright deposits prepared in the bell jar at one micron air pressure and on black deposits prepared in 13 mm of nitrogen. These deposits contained up to 14 mole percent tungsten oxides and showed increasing electrical resistivities when exposed to oxygen or air. The effect of various atmospheric gases on the electrical resistance of a bright deposit condensed at one micron pressure is shown in Fig. 1. The rate of deposition was 1.0 micrograms per square centimeter per minute, and the surface density was 34 micrograms per square centimeter. The resistance increased on initial exposure, then tapered off, and after several days became nearly constant. As shown in the figure, the initial rate of increase of resistance was much higher for oxygen than for air and was very small for carbon dioxide and nitrogen, the slopes of the curves in these latter cases being due probably to oxygen impurities and gas adsorption.

Similar results are shown in Fig. 2 for a gold smoke prepared in 13 mm of nitrogen at a rate of 9.6 micrograms per square centimeter minute with a surface density of 105 micrograms per square centimeter and containing about 1 mole percent tungsten oxides. The initial rate of increase of resistance was high upon exposure to oxygen and air, small upon exposure to nitrogen and carbon dioxide, as in the case of the bright deposits, and tapered off in a manner similar to the behavior of the bright deposits. The resistance of a black usually is the order of megohms, and it frequently rises to several thousand megohms on standing in air for several days.

The Effect of Oxygen Pressure on the Resistance

If a gold smoke containing tungsten oxides is exposed successively to higher oxygen pressures, the initial rates of increase in resistance can be shown to be a function of pressure. This is illustrated in Fig. 3 where $\log_{10}dR/dt$ is plotted against $\log_{10}P_{O2}$. This black was evaporated in 13 mm of nitrogen at a rate of 4.4 micrograms per square centimeter per minute. It had a thickness of 160 micrograms per square centimeter and contained about 2 mole percent tungsten oxides. The initial exposure of this deposit to oxygen resulted in an abnormally high resistance increase which fell off rapidly as the pressure was increased. At about 30-mm pressure, $\log_{10}dR/dt$ goes through a minimum, and above this pressure, it is a linear function of $\log_{10}P_{O2}$. This suggests that above 30 mm, the change in resistance of the gold smoke with respect to time may be represented by an equation of the form:

$$dR/dt = K(P_{O_2})^m, \qquad (2)$$

in which the constants K and m may be determined empirically. The low pressure anomaly, shown in an extreme case in Fig. 3, is more pronounced for blacks containing a smaller quantity of tungsten and having a relatively low resistance.

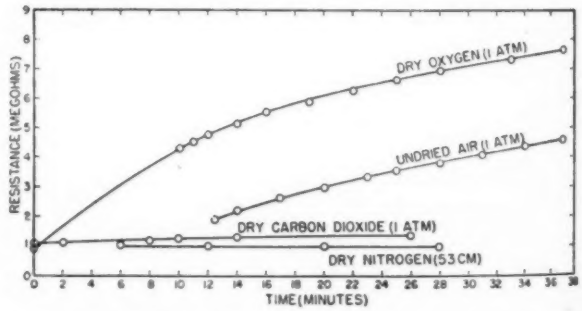


FIG. 2. The effect of atmospheric gases on the electrical resistance of a gold black condensed in 13 mm of nitrogen. Rate of deposition is 9.6 $\mu g/cm^2/min$. Thickness is 105 $\mu g/cm^2$. Sample contained 1 mole percent tungsten oxides.

The Electrical Cross Section of a Gold Black

The computation of electrical resistivity requires a measurement of the thickness in addition to the width and length of the sample. Blacks are usually flocculent loosely packed deposits, and the calculation of specimen thickness from a measurement of the weight of deposit per square centimeter depends on the density. However, what is needed is only an effective electrical thickness, i.e., the effective thickness of the material to the passage of electrons. Neglecting the comparatively small effects arising from varying contact resistances between the small particles for different samples, the assumption of bulk density (18.88 g/cm³) should be satisfactory for a calculation of resistivity, although this will not, of course, give the apparent physical thickness.

The Electrical Properties and Composition of Gold Smokes

For a number of small samples of blacks prepared under standardized evaporation conditions, the percentage of tungsten oxides was calculated from Eq. (1), and the initial electrical resistivity, ρ_i, before oxygen was admitted, and the constant K (in resistivity units) were calculated. Thus, in Fig. 4, $\log_{10}\rho_i$ is seen to rise rapidly from -3.23 ($\rho_i=6\times10^{-4}$) at 0.1 mole percent to 1.1 ($\rho_i=13$) at 2 mole percent tungsten oxides. Above 2 mole percent, $\log_{10}\rho_i$ rises much more slowly to a value of 1.97 ($\rho_i=93$) at 14 mole percent. The visual color of these samples varied from yellow at 0.1 mole percent to dark brown at 1 mole percent and became black at higher values. In like manner, the constants K and m in Eq. (2) are functions of the total tungsten oxide content. This is shown in Fig. 5. At the top is given $\log_{10}K$ which increased from -3.6 at 0.7 mole percent

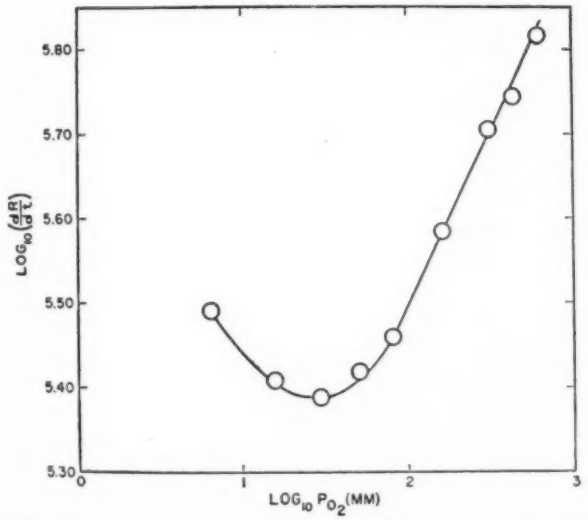


FIG. 3. The rate of change of electrical resistance as a function of oxygen pressure for a gold black condensed in 13 mm of nitrogen. Rate of deposition is 4.4 μg/cm²/min. Thickness is 160 μg/cm². Sample contained 2 mole percent tungsten oxides.

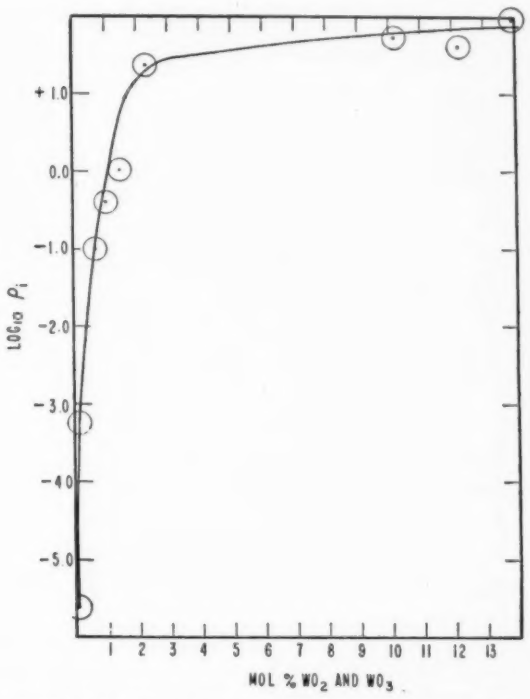


FIG. 4. The initial resistivity of various gold blacks as a function of the total mole percent tungsten dioxide and trioxide.

to -0.2 at 3 mole percent oxides and changed slowly above 3 mole percent until it reached 0.6 at 14 mole percent. The exponent m is also seen to rise (Fig. 5) from 0.18 at 0.5 mole percent and gradually flatten out with a value of 0.55 at 14 mole percent oxides.

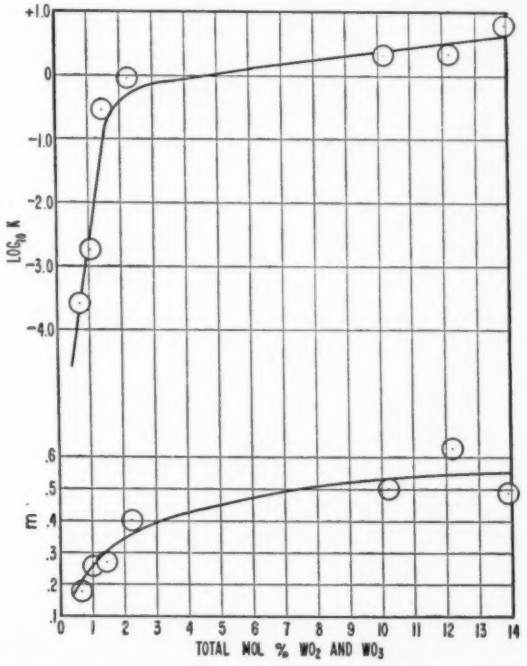


FIG. 5. The constants K and m (in equation $d\rho/dt=KPO_2^m$) for various total concentrations of tungsten dioxide and trioxide in gold blacks.

TABLE I. Electrical properties of evaporated tungsten trioxide and tungsten dioxide. ρ_i is the initial resistivity and m and K are constants in the equation: $d\rho/dt = K(P_{O_2})^m$.

Sample	Surface density $\mu g/cm^2$	ρ_i*	m	K
WO_2 evaporated at one micron from a quartz cup.	27	0.60	0.51	0.20
	30	0.21	0.47	0.18
	116	0.072	0.14	0.086
	162	0.0161	0.04	0.0078
WO_3 evaporated at one micron from a quartz cup.	573	2.7	0.39	2.3
Tungsten oxides formed by filament oxidation. $P_{N_2}=4.7$ mm. $P_{O_2}=0.5$ mm.	520	510	0.39	4.2

* For the calculation of electrical thickness, the density of WO_2 was taken as 12.11 g/cm³, and that of WO_3 was taken as 7.16 g/cm³. For the oxides produced by filament oxidation, the density was taken as 7.16 g/cm³.

Heating Gold Smokes

Heating gold smokes to 100 to 350°C causes a change in appearance from a black to a yellowish color accompanied by an increase in reflectivity,[6] which is, of course, undesirable when used as absorbing coatings in bolometers and thermocouples. On heating for 70 minutes at 350°C, one of our smokes showed a 4 percent loss in weight accompanied by a decrease in tungsten content and a change from a black to a yellow color. This effect is probably connected with the evaporation of the volatile tungsten trioxide and possibly a chemical change of other impurities.

Conclusions

In order to produce a gold black it is, thus, necessary to employ a gas atmosphere and to evaporate simultaneously a foreign substance such as tungsten trioxide with the gold. The best absorbers of radiation probably contain greater than 2 mole percent tungsten oxides. These blacks have resistances of the order of megohms. It is suggested that gold blacks, containing silver or copper, might be good radiation absorbers in the infrared and would possess a much lower stable resistance.

EVAPORATED TUNGSTEN OXIDES

Electrical Properties of Tungsten Oxides

Chemically prepared tungsten trioxide and dioxide, when evaporated in contact with a bare tungsten filament or from a quartz cup, and oxides, produced by tungsten filament oxidation, show, on exposure to oxygen, the phenomenon of increasing electrical resistivities which is very similar to that previously noted for gold smokes and films. In Table I are given some values of ρ_i, m, and K for a number of samples. The first four examples show that increasing the surface density of tungsten dioxide evaporated at one micron from a quartz cup causes ρ_i, m, and K to decrease. The values of these constants are of the same order of magnitude as those for gold smokes containing tungsten oxides. Tungsten trioxide, evaporated at one micron from a quartz cup, gave values for the initial resistivity and for K of 2.7-ohm cm and 2.3-ohm cm/min which are somewhat higher than those obtained for the dioxide. Tungsten oxides formed by filament oxidation in a mixture of 4.7 mm of nitrogen and 0.5 mm of oxygen gave high values of the initial resistivity and of K as shown. The value of m for both of these last two samples is nearly the same as the values obtained on tungsten dioxide and tungsten oxide containing gold smokes. In the calculation of the sample thickness for computation of resistivity, the density of tungsten dioxide was assumed equal to the bulk density (12.11 g/cm³). For tungsten trioxide and tungsten oxides formed by filament oxidation, the density was assumed to be that of tungsten trioxide (7.16 g/cm³).

Thermodynamic Properties of Tungsten Oxides

It is probable that the above observations can be explained by an oxidation of the conductive tungsten dioxide to the less conductive tungsten trioxide[11] according to the equation:

$$WO_2+\tfrac{1}{2}O_2 \rightleftharpoons WO_3, \qquad (3)$$

for which the equilibrium constant is 4×10^{41} at 300°K.[12] When tungsten trioxide is evaporated in vacuum from a quartz cup, considerable dissociation occurs by the reversal of reaction (3) in the gas phase above about 1400°K. The equilibrium constant of this reverse reaction is 1.9×10^{-5} at 1500°K and 2.6×10^{-3} at 2000°K.

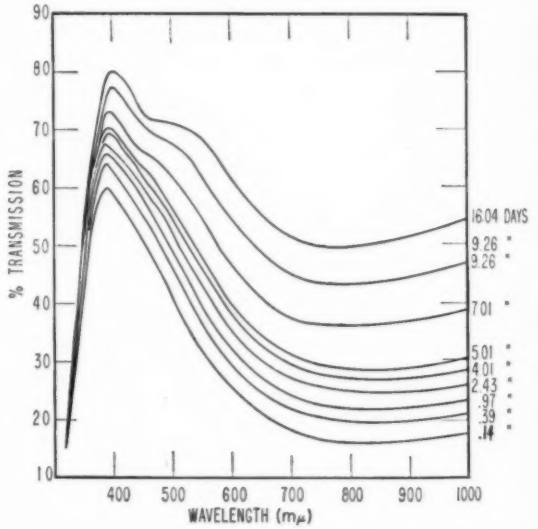


FIG. 6. The transmission of a vacuum evaporated tungsten dioxide film over a 16-day period. Rate of deposition is 5.4 $\mu g/cm^2$/min. Thickness is 162 $\mu g/cm^2$.

[11] Friederich, Z. Physik. **31**, 813 (1925).

[12] Thompson, *Total and Free Energy of Formation of the Oxides of 32 Metals*, New York City (1942). Seltz, Dunkerly, and DeWitt, J. Am. Chem. Soc. **65**, 600 (1943). Kozo Ueno, J. Chem. Soc. (Japan) **62**, 990 (1941).

When the trioxide is evaporated in vacuum in contact with metallic tungsten, the thermodynamic conditions are even more favorable and reduction proceeds according to the reaction:

$$2WO_3(s)+W(s)\rightleftharpoons 3WO_2(s), \qquad (4)$$

for which the equilibrium constant is 3.2×10^6 at 2000°K.[13]

Optical Transmission of Tungsten Oxides

Optical transmission measurements on evaporated films of tungsten dioxide support the electrical evidence for room temperature oxidation to the trioxide. Vacuum evaporated films of tungsten trioxide and dioxide both appear blue by transmitted and reflected light, adhere well to the glass collecting surface, and become more transparent on aging in air. The transmission measurements were made by means of a Beckman Spectrophotometer in the range from 320 mμ to 1000 mμ. The spectral transmission of a dioxide film, evaporated in vacuum at a rate of 5.4 micrograms per square centimeter per minute to a thickness of 162 micrograms per square centimeter, is shown in Fig. 6 for various times after preparation. The peak transmission occurs in the blue at 400 mμ, which is in agreement with its blue color. A broad absorption region occurs at about 800 mμ with an apparent narrower band developing at about 470 mμ. As shown, the transmission increased considerably over the 16-day period due to oxidation to the trioxide. During the same time, the electrical resistance increased from 1800 ohms to 2000 megohms.

Figure 7 shows the spectral transmission over the same wavelength region of a tungsten trioxide film vacuum evaporated from a quartz cup at 28 micrograms per square centimeter minute to a thickness of 570 micrograms per square centimeter. A broad absorption region occurs at about 900 mμ, and a transmission maximum is observed at 420 mμ. As can be seen qualitatively from Figs. 6 and 7, the trioxide is much more transparent than the dioxide. The trioxide showed small

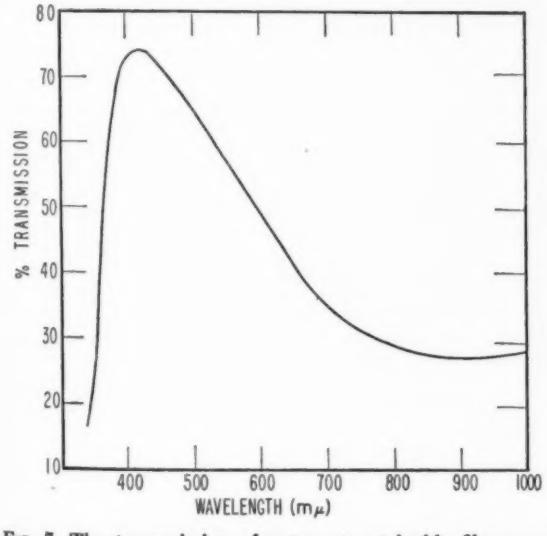


FIG. 7. The transmission of a tungsten trioxide film vacuum evaporated from a quartz cup. Rate of deposition is 28 μg/cm²/min. Thickness is 570 μg/cm².

[13] Schumb and Hartford, J. Am. Chem. Soc. 56, 2613 (1934).

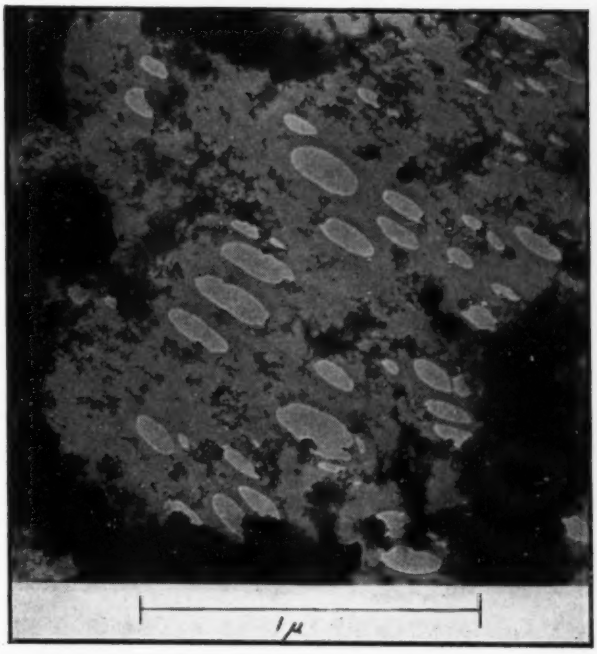


FIG. 8. Electron micrograph of a gold black evaporated in 13 mm of nitrogen on formvar. Particle size is 60A. Sample contains about 1 mole percent tungsten oxides.

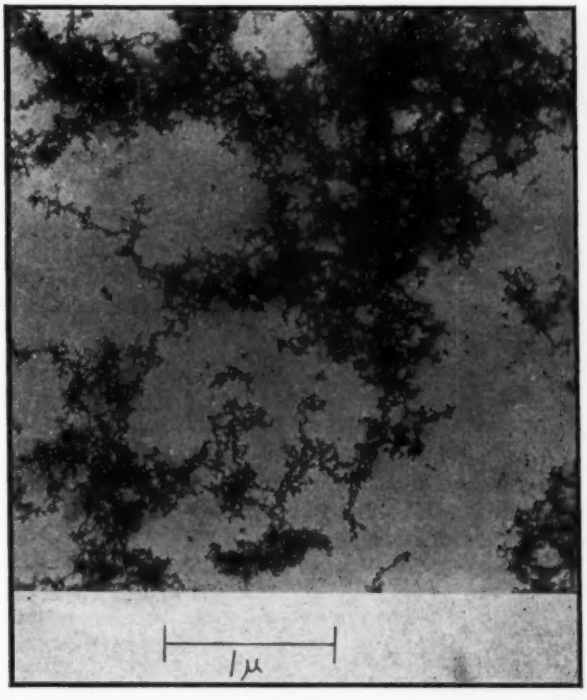


FIG. 9. Electron micrograph of tungsten trioxide evaporated in 10 mm of nitrogen.

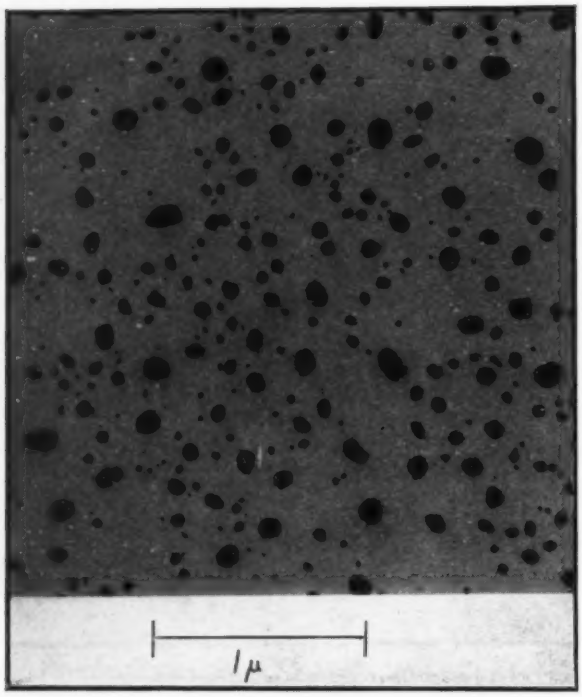


FIG. 10. Electron micrograph of a gold smoke formed in 1 mm of nitrogen and containing no tungsten oxides. Particles are typical of high vacuum evaporated gold.

increases in transmission and resistance on aging in air which can readily be explained by a partial thermal conversion to the dioxide during the evaporation process and subsequent oxidation.

ELECTRON MICROSCOPY

In Fig. 8 is shown an electron micrograph of a gold smoke which contained about 1 mole percent tungsten oxides. The specimen was made by condensing the smoke in 13 mm of nitrogen on a formvar membrane which developed holes in the process. The large agglomerates shown consist of many small particles with an average size of 60A.

In Fig. 9 is shown an electron micrograph of tungsten trioxide evaporated in 10 mm of nitrogen and, also, condensed on a Formvar membrane. The similarity in appearance with Fig. 8 is striking.

In Fig. 10 is shown an electron micrograph of a gold smoke condensed in 1 mm of purified nitrogen onto a formvar support and which contained less than 0.1-mole percent tungsten oxides. The particles formed irregular crystals, varied in size from less than 100A to 2000A, and were hardly distinguishable from high vacuum evaporated gold.[2] They were always obtained when the smoke contained less than about 0.5 mole percent tungsten oxides.

Electron micrographs of chemically prepared tungsten trioxide and dioxide films, deposited at one micron pressure, appeared structureless and identical with those formed by filament oxidation.[8] These materials can also be used for shadowing electron microscope specimens.

CONCLUSIONS

It is well known that impurity metal atoms cause a considerable increase in the resistivity of a metal, and it is not entirely surprising that traces of foreign oxides in gold should cause the tremendous increase shown in Fig. 4. It is difficult to predict the magnitude of the effect, since the physical state of the tungsten oxides in gold smoke is unknown. No particles typical of pure gold are found on electron microscope examination of smokes containing greater than 0.5 mole percent tungsten oxides. When both types of particles are present, the smoke has a low resistivity, is little affected by exposure to air, and has a dark brown appearance.

It is suggested that some of the aging effects, observed in the transmission of evaporated gold films, may be due to an oxidation of the tungsten dioxide probably present as an impurity.

The production requirements for a gold black are: evaporation of the gold in a gas atmosphere and the simultaneous evaporation of a foreign substance such as tungsten trioxide. It is not necessary to produce the tungsten oxides by oxidizing the tungsten filament; blacks are readily formed by evaporating in purified nitrogen a mixture of a few percent tungsten trioxide and gold powder. Both the nitrogen and the tungsten oxides (or other impurities) must be present; otherwise a yellow deposit results.‡

ACKNOWLEDGMENT

It is a pleasure to acknowledge the help of Mr. G. Gourlay, Mr. R. T. Seal, and Mr. S. Cress in performing the experimental work and of Mr. L. S. Birks for pertinent comments.

‡ The blacks prepared by Harris *et al.* (see reference 6) probably contained tungsten oxides in just the right concentrations to be good radiation absorbers. Harris, Jeffried, and Siegel [J. Chem. Phys. **18**, 261 (1950)] have also reported some gold blacks whose electrical resistance decreased as much as 70 percent after preparation and apparently did not increase on exposure to air. These deposits probably do not contain tungsten oxides but may contain small amounts of other metals such as silver and copper which could influence the optical and electrical properties.

Letters to the Editor

Condensation of Water Vapor in the Shock Tube Below 150°K

P. WEGENER AND G. LUNDQUIST
U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland
November 20, 1950

THE interest in condensation phenomena was revived in aeronautical laboratories with the advent of hypersonic wind tunnels in which air can be expanded below its "dew point" at high Mach numbers. The cloud-chamber tests by Cwilong,[1] showing absence of water vapor condensation for quick expansions to temperatures below 150°K, offered interesting new conjectures for those concerned with the parallel phenomenon of air condensation. Head's tests[2] in a supersonic tunnel operated with moist air seemed to verify Cwilong's observations. It was found at NOL,[3] however, that although supercoolings up to 100°C could be attained in a small wind tunnel also operated with moist air, condensation shocks could not be avoided.

To resolve this discrepancy, a 20×20-cm shock tube was used to produce expansions of moist air of accurately controlled supply conditions, where the resulting presence or absence of condensation can be detected by electronic methods. The compression chamber of the shock tube is filled with filtered air of controlled humidity at atmospheric pressure and room temperature. The expansion chamber, separated by a cellulose acetate diaphragm, is evacuated. When the diaphragm is punctured, it shatters, and a shock wave is propagated into the expansion chamber. The speed of this shock wave is measured and the flow speed behind it can be computed. The shock wave is followed by a contact surface traveling with the flow speed and separating the parts of the gas originally in the two chambers. The gas behind the contact surface is "cold." It has undergone an isentropic, unsteady expansion from its controlled initial conditions. The Mach number of the flow in this region is determined by measuring the angle of a shock wave produced by a 40° cone as shown in a schlieren picture taken with a 1 μsec exposure at a predetermined instant after the diaphragm burst (Fig. 1). With flow velocity computed from measured shock speed and Mach number known from the shock angle on the cone, the speed of sound can be computed and the temeperature derived. (The temperature and Mach number of the flow behind the contact surface cannot be reliably predicted, like flow velocity, from the theory, because of the erratic process of diaphragm breakage.[4])

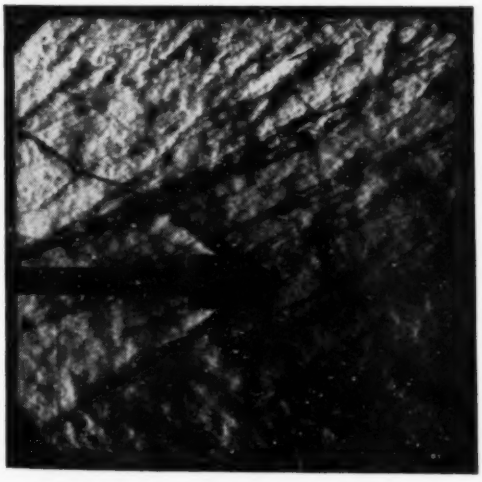


FIG. 1. Forty-degree cone in the shock tube. Mach number ~2.67.

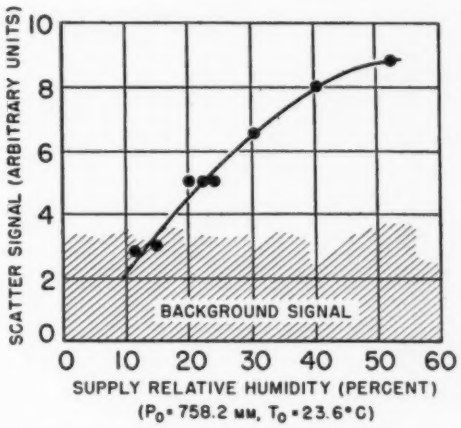


FIG. 2. Water vapor condensation at a temperature of 129° (±4°) K in the shock tube.

Condensation, if present in the shock tube flow, is detected by a specially developed sensitive electronic method in which the light scattered from droplets is observed and recorded.[5] Simultaneously with taking the picture, any light scattered and reflected at right angles to the parallel light beam of the spark schlieren system is picked up by a photo-multiplier tube and recorded on an oscilloscope screen. Figure 2 gives an example of the response for an expansion with varying initial humidity. The gas in this case has remained at the low temperature after its initial expansion approximately 4 msec before reaching the point of observation. The identical experiment was repeated several times with dry air to obtain the magnitude of the background signal produced by light scattered from dust, diaphragm particles, and the shock tube walls.

It can be seen that condensation of water vapor was detected at a temperature of 130°K, 20 degrees below the threshold of 150°K set by Cwilong's tests. It was found that a sensitive photo-multiplier tube is necessary for the detection of light scattered from low temperature fogs of short duration where visual observation gives negative answers. It is believed that, in so far as the rapidity of the expansion affects the occurrence of condensation, the water vapor in the shock tube case had considerably less time available for the formation of a liquid phase than it would have had in a cloud chamber.

[1] B. M. Cwilong, Proc. Roy. Soc. (London) **A190** (1947).
[2] R. Head, "Investigation of spontaneous condensation phenomena," GALCIT Dissertation (1949).
[3] P. Wegener, Phys. Rev. **76**, 883 (1949).
[4] F. W. Geiger and C. W. Mautz, Phys. Rev. **79**, 230 (1950).
[5] G. Lundquist, "The shock tube as an instrument for the investigation of transonic and supersonic flow patterns," Eng. Res. Inst., University of Michigan, Project M720-4 (1949).

Observation of Initial Flow in a Fluid Obeying Darcy's Law by Radioactive-Tracer Techniques

F. T. ROGERS, JR., AND L. E. SCHILBERG
U. S. Naval Ordnance Test Station, China Lake, California
December 4, 1950

IN 1947 H. L. Morrison and others reported[1] the observation of initial convective flow in fluids obeying Darcy's law in porous media, using visual methods[2] limited to regions of fluid adjacent to vessel walls. The migration of dissolved dyes, the essential simultaneity of initial flow at all depths (adjacent to vessel walls), and an associated "discontinuity" in heat transfer,[3] all suggested that the flow occurred quite generally through the porous medium, though, to be sure, no direct evidence had been obtained to this effect.

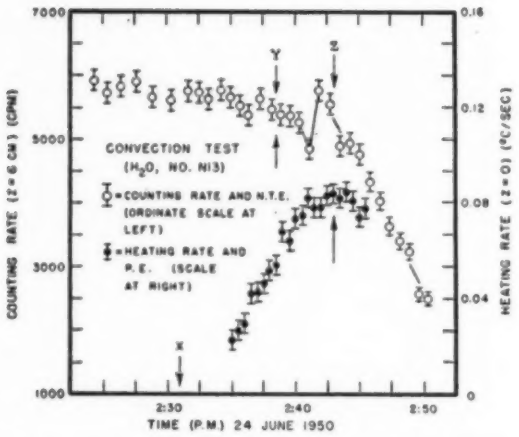


FIG. 1. Radioactive, thermal, and visual observations relative to the onset of convective flow of water in a sand layer 10.3 cm thick and having a flow permeability of the order of 10^3 darcys. Heating began at X; only diffusive motions were visible at times such as Y; and convective flow was visible at Z.

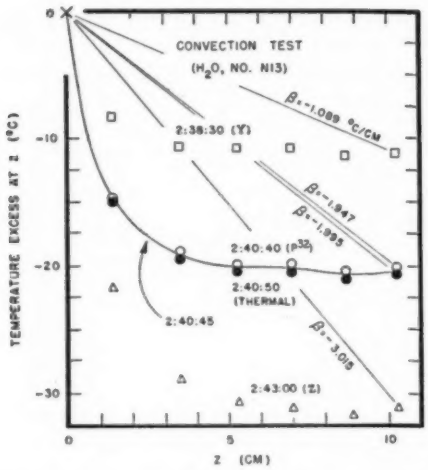


FIG. 2. Temperatures relative to bottom temperature, and corresponding mean thermal gradients, at various times identified in Fig. 1. This graph illustrates the relation among mean gradients as indicated by the several methods of observation, and the functional form $f(t)\cdot g(t)$ for the temperature excess.

Following up a suggestion advanced at the Baton Rouge meeting,[2] we have sought to identify in more detail the properties of initial flow by means of changes in counting rate of a Geiger-Müller counter placed near a locally radioactive inoculation of the fluid. Essentially, we have repeated some of the tests described in reference 2, with a small medical-type counter immersed in the fluid, and employing improved thermocouple thermometry. Prior to this, limited tests had been made in collaboration with Dr. H. D. Bruner of the Oak Ridge Institute for Nuclear Studies (November, 1949) using the 1.17-Mev gamma-radiation from Zn 65 and a counter external to the fluid mass; and these had borne out calculations to the effect that this arrangement would be insensitive at so small a scale of dimensions.

Figures 1 and 2 show the data obtained in one such test, using the 1.71-Mev (maximum energy) beta-radiation from P 32. After micropipette inoculation with about 0.1-microcurie activity of dilute $H_3P^{32}O_4$ solution[4] near the counter's sensitive volume, counting was started (upper curve in Fig. 1); and when a stabilized counting rate indicated settling to have ended, the heater was turned on (time X in Fig. 1). Visual observations of dissolved methyl orange dye were made continually; at times such as Y in Fig. 1, only diffusive motion was discernible, while at Z the abrupt beginning of convective flow was seen. Bottom heating rates were read from the thermograph after the test was concluded (lower curve in Fig. 1). Three indications of the onset of flow were thus obtained: in this test, radioactive at time 2:40:40, thermal at 2:40:50, and visual at 2:43:00. Figure 2 shows the temperature excess (above bottom temperature) at each of these times, as a function of the distance (z) above the bottom of the vessel, together with the corresponding mean thermal gradients (β).

The test reported in Figs. 1 and 2, together with other like tests which we have carried out, constitutes a body of observations in support of the previous supposition[1-3] that initial flows observed visually do indeed extend throughout the porous medium. Further than this, however, the present experiments indicate that initial flow begins a little sooner (at smaller absolute values of the mean gradient, β) than would be judged by the visual method. They also indicate a functional form $f(t)\cdot g(z)$ for the temperature excess, from which one easily obtains[5] values of thermal diffusivity which are somewhat higher than had been used in references 2 and 3. Both the smaller gradients and the larger diffusivities operate to increase by a factor of about 3 or 4, that disparity between experiment and theory which was reported in reference 2. On the other hand, the form $f(t)\cdot g(z)$, which has now been found, considerably simplifies our attempts to improve the extended theory[6] by an approach related to that of H. Jeffreys. An interesting by-product of this investigation is the rather high sensitivity of the radioisotope technique as used; we attribute this sensitivity to the fact that when flow begins, geometry, absorption, and the nature of the beta-particle spectrum all operate in the same sense in altering the counting rate.

[1] H. L. Morrison, Phys. Rev., **71**, 834 (1947), and subsequent papers.
[2] Morrison, Rogers, and Hoston, J. Appl. Phys. **20**, 1027 (1949).
[3] F. T. Rogers, Jr., and H. L. Morrison, Contributed Paper No. T1, Sixteenth Annual Meeting of the Southeastern Section of the American Physical Society, April 7–8, 1950.
[4] The active material was obtained from the Oak Ridge National Laboratory under authorization number 5849 of the Isotopes Branch, ORO, U. S. AEC, Oak Ridge, Tennessee.
[5] See E. B. Penrod, J. Appl. Phys. **21**, 425 (1950).
[6] F. T. Rogers, Jr., and H. L. Morrison, J. Appl. Phys. 21, 1177 (1950).

A Dynamic Method for Measuring the Specific Heat of Metals

ALAN M. NATHAN
Research Division, New York University, University Heights, New York*
September 18, 1950

IN this letter is described a dynamic method of measuring the specific heat curves of metals under conditions of rapidly increasing temperature (1000°C per second) together with some results obtained with it. The method is used for a preliminary study of the specific heat curves of low carbon steel at the phase transformation points.

Ordinary lead storage batteries were chosen as the most convenient source of electrical power capable of producing rates of rise of temperature from 1000°C per second to 1000°C per minute in a small steel specimen. By suitable variation of the number of cells and hence the applied voltage, it is quite possible to secure any desired rate of heating, since such batteries are capable of supplying currents up to several hundred amperes for short periods. The steel specimen to be heated is in the form of a cylindrical rod 20 cm in length and 0.3 cm in diameter. When the specimen is heated by an electric current, i, the heat liberated per gram of material during the time dt is

$$dQ = ei\,dt/J\,M,$$

where e is the voltage across the central portion under test, J is Joule's equivalent, and M the mass, provided one neglects heat losses due to radiation from the surface of the heated rod, which are negligible compared to the total heat input at high rates of heating. The heating effect of the current is essentially uniform throughout the material owing to the large length-to-diameter

ratio of the rod, while the loss of heat due to heat conduction to the end fittings is very slight during the short time duration of a run.

The quantity e is measured by recording the voltage drop across the central portion of the specimen under test using two pick-up leads (see Fig. 1). The heating current, i, is measured by recording the voltage drop across a low value ohmic resistor in series with the specimen. These two voltages are combined on one trace of the recording dual beam scope by means of a dc coupled electronic switch. Temperature of the specimen is measured by a platinum-platinum 13 percent rhodium thermocouple with wires 0.002 in. in diameter and its junction spot welded to the surface of the specimen at its center. The diameter of the spot weld is approximately 0.004 in. Tests show that the thermal lag of this temperature measuring system is negligible at the rates of heating used. The temperature signal, T, is converted to a modulated ten-kilocycle carrier by a dc to ac diode bridge modulator and displayed on the second oscilloscope trace.

The specific heat at constant pressure, C_p, is defined as $C_p = dQ/dT$; therefore,

$$C_p = \frac{ei}{JM\, dT/dt}.$$

To compute C_p as a function of temperature, then, the three quantities e, i, and T must be recorded during a heating run. A

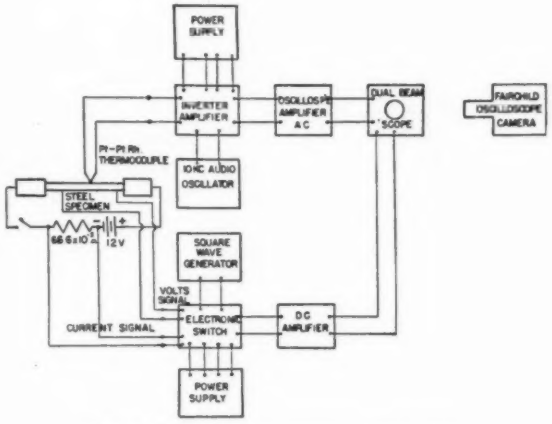


FIG. 1. Specific heat measurement apparatus.

graphical differentiation of the temperature record from a 5-in. oscilloscope screen showed too much scatter for a precise determination of specific heats. Nevertheless, an order of magnitude study was made and the values of the specific heat obtained agreed well with the statically measured values. To eliminate the errors inherent in the graphical differentiation process, it was thought that electronic differentiation of the temperature signal before recording would give an accurate determination of the term dT/dt.

Since the specific heat of the specimen is inversely proportional to the time derivative of temperature, any sharp changes in the slope of the temperature *versus* time record indicate sharp changes in the value of the specific heat at that temperature. The power input is fairly constant with time and does not undergo sudden discontinuities, so that peaks in the specific heat curve will show up as breaks in the slope of the temperature *versus* time trace. By measuring the temperature at which these breaks occur, comparisons can be made between the occurrence and location of peaks during rapid heating and very slow or equilibrium conditions.

This procedure has been carried out for a series of heating runs of different durations, and slight dips in the temperature records were observed on almost all traces. In order to measure the temperatures at which these points occurred, the negatives of the oscilloscope temperature trace were projected as magnified images onto the screen of an optical comparator. With the use of a straightedge and the micrometer adjustments of the comparator

TABLE I.

Duration of run in seconds	Temperature at dip uncorrected			Average temp. at dip	Deviation
1.5	773°C	768°C	775°C	772°C	2.6°C
2.5	772	771	768	770.3	0.9
3.0	764	761		762.5	6.9
5.5	771	774	773	772.7	3.3
	Average temperature at dip			769.4°C	±3.4°C av. dev.

table it was possible to measure the location of these dips with reproducible accuracy. The data obtained showed the points to correspond to the specific heat peak occurring at the Curie point, 769°C, within the limits of accuracy of the measurement. However, peaks in the specific heat curve that normally appear in the neighborhood of 720°C under equilibrium conditions owing to an allotropic phase change in steel were not observed during rapid heating.

Observations.—Three readings were taken on each of the four useful records obtained; the temperature at which the dip occurs being measured. Presented in Table I are the results of these observations.

* The work described in this paper was done with the support of the ONR, Department of the Navy, and the Office of Air Research, Department of the Air Force, under Contract N6ori-11, Task Ordnance 2, as part of Project Squid.

Random Noise in Dielectric Materials

J. H. MASON
The British Electrical and Allied Industries Research Association, Greenford, Middlesex, England

October 6, 1950

THE current fluctuations observed when thin films of polar plastic materials are subjected to dc potentials of 15 to 300 v per mil have been attributed by R. F. Boyer[1] to the motion of groups of ions in the direction of the field, or in the reverse direction, to relieve the concentration gradients set up by the preceding forward motion. An alternative and comparatively simple explanation which obviates the difficulties encountered by the theory advanced by the author is that minute surface discharges occur between the metal electrodes and the dielectric surfaces. At a critical discharge inception stress discharges may occur between many small areas of the surface, not simultaneously, but during several seconds or minutes, depending on the nature of the dielectric, the thickness of the air film between the electrodes and solid dielectric, and the applied stress. These discharges continue with decreasing frequency until over the whole surface the voltage between the dielectric and the electrode has decreased to the extinction voltage. Thereafter, discharges recur only as a result of charge leakage, dependent on the relaxation time constants of the material; if the applied stress is removed, some back discharges occur, and reversal of the stress results in numerous reverse discharges until they are again inhibited by the accumulation of surface charges. With ac stress continuous discharges would be observed.

With dry samples the discharge inception stress was apparently some 120 kv/cm across the sample, and assuming the permittivity was about 3, the stress in the micro-air-gap would be some 360 kv/cm; at this stress discharges would occur in an air gap of 10^{-3} cm, according to the Paschen curve. At high relative humidity, however, noise was observed at much lower stresses (e.g., about 12 kv/cm across the sample, in the case of Nylon at 86 percent relative humidity); and further explanation is necessary. There is some evidence that with microgaps at atmospheric pressure the breakdown characteristic does not follow Paschen's law, and that glow discharge may occur at comparatively low stresses.[2,3] The presence of moisture may increase the permittivity of the

dielectric, and therefore the stress in the air gap, and may also reduce the stress at which glow discharges occur.

Gross[4,5] has recently shown that the permanent "homo" charge (of the same sign as the electrode) on electrets arises from a transfer of charge, frequently by micro-gas-discharges, from the electrode to the dielectric surface. Discharges between a disk of polystyrene and an adjacent electrode, at 0 percent relative humidity commenced at about 80 kv/cm, in the air gap.[5]

The writer has shown by electrophotography[6,7] that individual discharges between dielectric surfaces involve only a small area, but that with continued application of the stress many separate areas may discharge until eventually the whole area is discharged.

The variation in magnitude of the noise observed with different dielectrics and with varying humidity is readily explained by the variation in surface and volume conductivity, which control the area which discharges as a single unit, and also the time constant of charge leakage.

Critical tests to determine the origin of the noise observed by Boyer would be as follows: (a) observe whether the noise is eliminated if evaporated silver or colloidal carbon electrodes are applied directly to the plastic materials, so that complete and uniform contact with the metal electrodes is obtained; (b) observe the polarity of any residual charge on the dielectric surface after reducing the electrode polarity to zero, using the technique of the demountable capacitor.[5]

[1] R. F. Boyer, J. Appl. Phys. **21**, 469 (1950).
[2] G. M. Hobbs, Phil. Mag. **10**, 617 (1905).
[3] R. W. Jones and W. S. Huxford, Trans. Illinois Acad. Sci. **34**, No. 2 (December, 1941).
[4] B. Gross, J. Chem. Phys. (to be published).
[5] B. Gross, E.R.A. Report Ref. L/T245, 1950 (awaiting publication in the *British Journal of Applied Physics*).
[6] J. H. Mason, Nature **164**, 451 (1949).
[7] J. H. Mason, E.R.A. Reports Refs. L/T210, 235 (awaiting publication in the Proc. I.E.E., London).

A Cylindrical Cavity Filled with a dc Discharge

GEORGE W. STUART, JR., AND PHILIP ROSEN
Rensselaer Polytechnic Institute, Troy, New York
October 5, 1950

IF one considers a cylindrical cavity with perfectly conducting walls that is filled with a dc discharge, the solutions to the boundary value problem for the circularly symmetric TM modes yield for the resonant frequency: $\omega=\Lambda[(1/\mu_0\epsilon)-(\sigma/2\epsilon\Lambda)^2]^{\frac{1}{2}}$. $\Lambda^2=(r_{0s}/a)^2+(l\pi/L)^2$. The gaseous discharge is treated as a lossy dielectric of permittivity ϵ and conductivity σ; r_{0s} is the sth root of J_0, l is an integer, and a and L are the radius and length of the cavity, respectively. The Q of the cavity is found to be $Q=c\Lambda\epsilon/\sigma$. If wall losses are considered, then

$$\frac{1}{Q}=\frac{1}{c\Lambda\epsilon/\sigma}+\frac{1}{Q_0},$$

where Q_0 is the Q with a vacuum in the cavity.

For a low pressure discharge with ω in the microwave region, we essentially have free electrons and $\epsilon=\epsilon_0(1-e^2N/m\omega^2\epsilon_0)=\epsilon_0(1-\omega_p^2/\omega^2)$, where ω_p is the "plasma resonant frequency."[1] The value of σ for this condition is very small. When $\sigma=0$ we find: $\omega^2=\omega_0^2+\omega_p^2$, where ω_0 is the resonant frequency with the vacuum.

Study of Eq. 1 shows that the cavity will have an imaginary frequency for negative ϵ when $\sigma=0$. However, for a small but finite σ the cavity will pass to the non-oscillatory condition when ϵ/ϵ_0 is a small positive quantity. This behavior is similar to that of the "coax" with the central conductor as the discharge, because the propagation characteristics change markedly at a certain value of the permittivity when σ is small.[2] For the cavity this occurs at $\epsilon=0$, and at $\epsilon/\epsilon_0=-1$ for the "coax."

It has been pointed out by Kerr[3] that such devices can be used as microwave switches or as frequency modulators.

[1] Eccles, Proc. Roy. Soc. (London) **87A**, 79 (1912).
[2] P. Rosen, J. Appl. Phys. **20**, 868 (1949).
[3] Kerr, Brown, and Kern, Phys. Rev. **71**, 480 (1947).

Erratum: Original Authorship of Formulas Given in Clark and Deutsch's Paper

[J. Appl. Phys. **21**, 1195 (1950)]

M. REINER
Institute of Technology, Haifa, Israel
and
G. W. SCOTT BLAIR
National Institute for Research in Dairying, University of Reading, England

Our Letter to the Editor in the November, 1950, issue was inadvertently published without the following references:

Hersey, J. Rheology **3**, 196 (1932).
M. Mooney, J. Rheology **2**, 210 (1931).
B. Rabinowitsch, Z. phys. Chem. **145A**, 1 (1929).
M. Reiner, *Deformation and Flow* (H. K. Lewis, London, 1949), p. 149.
G. W. Scott Blair, *Introduction to Industrial Rheology* (J. & A. Churchill, London, 1938), pp. 37–38, 79.